STRUCTURAL STABILITY, DYNAMICS AND UNFOLDING OF 7-TRANSMEMBRANE HELICAL RECEPTORS

By James Mitchell

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Medical Sciences

University of Warwick, Warwick Medical School

April 2018

CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Introduction

Rhodopsin and vision

Establishment of pSRII as an experimental system for unfolding

Further development of rhodopsin as a model system for NMR studies of folding and misfolding

Rhodopsin misfolding

ACKNOWLEDGEMENTS

Thanks first to my supervisors. Judith Klein-Seetharaman's eye for detail and ability to squeeze all possible data from experiments are examples I will try to follow. Her incredible work ethic even after transatlantic flights is something I can only hope to emulate. Narinder Sanghera persuaded me to do this in the first place and has supported me through it with her biochemistry expertise and calm presence.

The advice of my friends and colleagues at Warwick was as valuable as the board games and trips to the pub that kept me going. Thanks for all your help. Special thanks to Joan Planas-Iglesias for not letting me get too lazy with my code.

My other friends were a great source of board games and pub trips, but their scientific advice was of no use whatsoever.

Greatest thanks to my greatest team mate, April, for whom the word girlfriend is woefully inadequate, and the word partner is too weird*. Both of our families have been an amazing base of support. I love you all.

I couldn't have done it without any of you.

*Between submitting my thesis and my corrections the problem has been resolved by getting engaged.

COLLABORATION DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Data provided by Yi Lei Tan and Daniel Nietlispach (University of Cambridge)
    - Some UV/Vis unfolding time courses of pSRII
    - UV/Vis refolding time courses of pSRII
    - Fluorescence unfolding time courses of pSRII
    - Circular dichroism unfolding time courses of pSRII
    - NMR spectra of pSRII
- NMR spectra of mammalian rhodopsin collected with the help of Yi Lei Tan and Daniel Nietlispach
- NMR peaks assigned by Yi Lei Tan, Daniel Nietlispach and Mark Bostock
- PNGase F treated gels from Daniel Mann

Parts of this thesis have been published by the author:

James Mitchell, Yi Lei Tan, Arpana Dutta, Daniel Nietlispach and Judith Klein-Seetharaman (2018) Rhodopsin and Sensory Rhodopsin II – Stability and Characterization of Unfolded Structures. Encyclopedia of Biophysics, Gordon C. K. Roberts (Ed.). In press

Iannaccone, A., Balem, F., Mitchell, J., Man, D., Dhiman, H.K., Yanamala, N., Ollesch, J., Tirupula, K., Planas-Iglesias, J., Waseem, N., Jennings, B.J., Gerwert, K., Bhattacharya, S.S. and Klein-Seetharaman, J. (2018) Comparison of the Molecular Properties and Retinitis Pigmentosa Disease Phenotypes of P23H and N15S Amino Acid Replacements in Rhodopsin. Plos One. Under review.

James Mitchell, Yi Lei Tan, Naveena Yanamala, Fernanda Balem, Eric E. Gardner, Kalyan C. Tirupula, Mudi Sheeves, Daniel Nietlispach and Judith Klein-Seetharaman (2018) Structural and functional consequences of the weak binding of chlorin e6 to rhodopsin. In preparation.

Yi Lei Tan, James Mitchell, Judith Klein-Seetharaman, Daniel Nietlispach (2018) Characterisation of SDS-Unfolded States of Sensory Rhodopsin II. In preparation.

ABSTRACT

The dynamics of membrane proteins is an understudied area due to the difficulties in production and manipulation of samples. Interactions of the G-protein coupled receptor, rhodopsin, with an allosteric reagent were systematically investigated through absorbance spectroscopy. Folding mechanisms in particular are not understood for membrane proteins, due to the fact that even partial unfolding often leads to aggregation. There are few model systems for exploring membrane protein folding, but methods for obtaining large quantities of non-aggregating unfolded states of rhodopsin have previously been established. In this thesis the previous work establishing rhodopsin as a model system was extended with the use of isotope labelled methionines for NMR spectroscopy of unfolded states. The misfolding rhodopsin mutant, P23H, was compared with wild-type using this system, and found to have significant differences. The same mutant was also found to be destabilised by normally benign cysteine labelling, which has implications for concurrent research. The first membrane protein to have been fully unfolded and refolded is bacteriorhodopsin, to which sensory rhodopsin II (pSRII) is an analogue. Computational work suggests that pSRII has a folding mechanism somewhat like both bacteriorhodopsin and rhodopsin. The kinetics of unfolding and refolding pSRII were established using experiments with low sample requirements, and thoroughly analysed. Several analytical methods were applied to find underlying processes in unfolded states, revealed through NMR spectroscopy.

ABBREVIATIONS

3S8U	3% SDS, 8M Urea
A8-35	Amphipathic polymer with mass ~8kDa and 35% derivatized backbone
ADRP	Autosomal Dominant Retinitis Pigmentosa
AIC	Akaike's "An information criterion"
AM-2	Trimethyl amine N-oxide based tandem amphiphile
APS	Ammonium persulphate
A. tittmanni	*Aristostomias tittmanni*
bR	Bacteriorhodopsin
B. taurus	*Bos taurus*
c7-DHPC	1,2-diheptanoyl-sn-glycero-3-phosphocholine
CD	Circular dichroism
Ce6	chlorin e6
cGMP	cyclic guanosine monophosphate
csv	Comma separated values
DET	Dexter energy transfer
DM	n-dodecyl β-d-maltoside
DMEM	Dulbecco's modified Eagle medium
DMSO	Dimethyl sulphoxide
EC	Extracellular
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
FBS	Foetal bovine serum
FIRST	Floppy inclusions and rigid substructure topography
FRET	Förster resonance energy transfer

G418	Geneticin G418
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GPCR	G protein coupled receptor
GTP	Guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMQC	Heteronuclear multiple quantum coherence NMR spectroscopy
HSQC	Heteronuclear single quantum coherence NMR spectroscopy
Hz	Hertz
LM	The Levenberg-Marquardt algorithm for NLS fitting
LMNG	Lauryl maltose neopentyl glycol
MES	2-(N-morpholino)ethanesulphonic acid
Meta II/III	Metarhodopsin II/III
M. niger	*Malacosteus niger*
MP	Membrane protein
MR	Mammalian rhodopsin
MRE	Molar residue ellipticity
MSA	Multiple sequence alignment
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
MW	Molecular weight
NaPi	Sodium phosphate buffer, pH 6
NLS	Nonlinear least squares fitting
NMR	Nuclear magnetic resonance
PBS	Phosphate buffered saline, pH 7.4

PC	Principal component
PCA	Principal components analysis
PDT	Photodynamic therapy
ppca	Probabilistic PCA
ppm	parts per million
PDS	4,4'-dipyridyl disulphide
PRE	Paramagnetic relaxation enhancement
pSRII	*Natronamonas pharaonis* sensory rhodopsin II
ROS	Rod outer segment
RP	Retinitis pigmentosa
rpm	revolutions per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SOC	Super optimal medium with catabolite repression
TBS	Tris buffered saline
TBS-T	TBS/Tween 20

CHAPTER 1: INTRODUCTION

1.1: Background

1.1.1: Rhodopsin and vision

There is a great variety in eye structure across the animals. It was thought that this variety came from multiple evolutionary origins, but similarities in the underlying transcription factors for eye development suggest a single, very ancient origin (Nilsson and Arendt, 2008). Some animals have "eyespots", which only comprise a light-sensitive patch of tissue , though more complex structures have some form of image-resolving structure, from the simple pinhole eyes of the nautilus to the lens apparatus of mammalian vision (the evolution of these different eye structures is reviewed in Land and Fernald, 1992). In these eyes, light passes through the lens, the aperture and focus of which may be controlled by surrounding muscles. These can be adjusted to form a coherent image on the light sensitive retina, which has a diverse population of cells responsible for the detection of light and early processing of vision (Jeon et al., 1998). There are other cell types and structures within the eye responsible for the maintenance of the essential structures, and some animals also have adaptations for particular visual purposes, including especially low-light conditions, as outlined in section 1.1.1.2.

The cells responsible for vision in vertebrates are the rod and cone cells of the retina. These cells have stacks of membrane discs in their outer segment containing high concentrations of opsins, a family of proteins which includes the mammalian cone and rod opsins responsible for daylight colour (photopic), and night-time (scotopic) vision respectively in mammals. Photoreceptor cells are specialized neurons, and have an inner segment providing energy to maintain a negative membrane potential, a nucleus, and an axon releasing glutamate (Lodish, 2008).

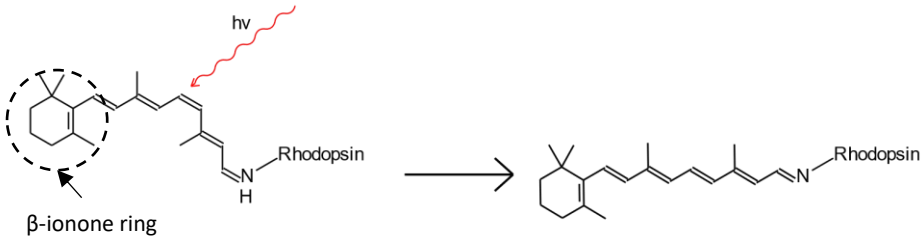


Figure 1.1: Photoisomerization of retinal. A photon is absorbed by 11-*cis* retinal, causing isomerization of the 11-12 bond from *cis* to *trans*.

1.1.1.1: Visual transduction

The first event in the photocycle of rhodopsin, the opsin found in rods, is the isomerisation of 11-*cis*-retinal to all-*trans*-retinal by a photon (Figure 1.1). The retinal molecule is covalently linked to an opsin protein via a protonated Schiff base bond to a Lysine (K296 in the case of the dim-light rod photoreceptor, rhodopsin). The first state in the photocycle of rhodopsin is photorhodopsin. This excited state thermally decays to bathorhodopsin, in which the protein structure is mostly the same, but the retinal has isomerised, leading to a highly strained retinal chain. This is exacerbated by the β-ionone ring being held in place. Though the protein's structure is mostly the same as the ground state, small changes have started in helix 3, and to the side chain of F212 in helix 5 (Nakamichi and Okada, 2006a). The retinal continues to relax through a blue-shifted intermediate to lumirhodopsin. In this state, the retinal has mostly reached a relaxed conformation, though there is some twisting in the chain. This relaxation moves the energy taken in by absorbing the photon from retinal to conformational changes in the protein. The ring is displaced towards helices 3 and 4, causing some local distortion to helix 3. The overall effects on the protein to this point are minor, mostly weakening the interactions between helices (Nakamichi and Okada, 2006b). The next state, metarhodopsin I, is very similar, the major change being to the Schiff base. In the ground state, and the preceding intermediates, the Schiff base is kept protonated by the counterions provided by E113 and E181, with the greatest effect coming from E113. In the metarhodopsin I state, the balance of this shifts to E181. The largest conformational changes to the protein occur in the transition to metarhodopsin II (Meta II). There are major changes to an internal hydrogen bond network (Vogel et al., 2008), and the Schiff base is deprotonated. The retinal binding pocket is opened by gaps forming between helices. Helices 1 and 7 separate, and further opening occurs through tilting of helix 6 and rotation of helix 5 (Choe et al., 2011). The transition from the ground state to Meta II takes ~7 milliseconds.

When rhodopsin is in the Meta II state, the change in the tilt of helix 6 and the rotation of helix 5 provide an opening, allowing its cytoplasmic domain to activate G proteins, passing the signal across the membrane. The G protein, transducin, is a heterotrimer of α, β and γ subunits. When inactive, the complex binds guanosine diphosphate (GDP), but the conformational change induced by rhodopsin allows GDP to be exchanged for guanosine triphosphate (GTP). This exchange causes the α subunit to dissociate, though β and γ remain tightly associated and remain as one subunit. The βγ subunit is capable of some signalling, though this is of little relevance to visual transduction. The α subunit goes on to activate phosphodiesterase, which breaks down cyclic guanosine monophosphate (cGMP) to

guanosine monophosphate (GMP). The lower concentration of cGMP results in less opening of the cell's cGMP-gated cation channels, and the membrane potential is hyperpolarised. This reduces the release of glutamate, thereby lifting its inhibitory effects on other neurons in the retina, allowing the propagation of the signal. A schematic of the process is shown in Figure 1.2. The processing of signals from rods and cones is complex and takes place both in the retina (Dowling, 1999) and in the visual cortex (Maier, 2013).

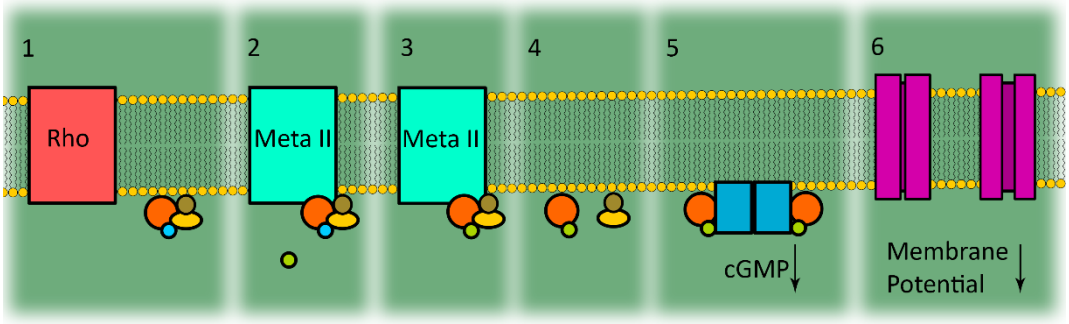


Figure 1.2: Visual transduction. 1) The retinal of rhodopsin is photoisomerized as in Figure 1.1 and rhodopsin undergoes conformational changes to metarhodopsin II. **2)** Metarhodopsin II is able to activate transducin. **3)** The GDP bound to inactive transducin is exchanged for GTP. **4)** Active transducin splits into its α and βγ subunits. **5)** The α subunit activates phosphodiesterase, decreasing cGMP concentrations in the cell. **6)** The decrease in cellular cGMP closes the cGMP gated cation channels, causing hyperpolarisation of the cell.

After activation of rhodopsin, a dedicated rhodopsin kinase phosphorylates rhodopsin and the protein arrestin binds. This stops rhodopsin signalling. The bound retinal is released, and Meta II decays. There are several states like metarhodopsin III (Meta III) and another storage states to which it can decay, but to continue the photocycle both retinal and rhodopsin are modified to yield 11-*cis*-retinal and opsin capable of binding it once again (Heck et al., 2003).

For human and bovine rhodopsin, the absorbance maximum is at 498nm, making it most sensitive to green light. It is the basis of scotopic, or dim-light vision. Other, less stable and less sensitive opsins are responsible for colour vision. All opsin functions are based on the same retinal cofactor for light detection, but the influence of different amino acids upon the retinal tune these to be most sensitive to different wavelengths, a mechanism referred to as the "opsin shift" (Bravaya et al., 2007). The different environments of animals results in the evolution of differently sensitive opsins, and other specializations in their visual apparatus.

1.1.1.2: Adaptations in vision of deep-sea fish

Fish in shallow waters tend to have vision not too dissimilar to terrestrial animals, often possessing both rods and cones and detecting light from a broad spectrum (Bowmaker, 1995). However, with greater depth, adaptations for specialised vision become common, as light becomes less available with greater depth and bioluminescence from ocean creatures

takes over as eventually the only light source. Light is attenuated by the water, and is filtered to progressively narrower bandwidths (Arst et al., 1997; Smith and Baker, 1981). From 200 to 1000 metres, downwelling light goes from being a twilight to a negligible influence (Warrant, 2000). Fishes in this region, the mesopelagic zone, are adapted to the almost exclusively blue-green space light (i.e. light from the sun or moon) in sometimes surprising ways. The visual pigments are commonly tuned to the downwelling light, around 470 nm. A strategy of many prey organisms is to use bioluminescence as camouflage, disguising their outline from deeper predators by counterillumination (Johnsen et al., 2004).

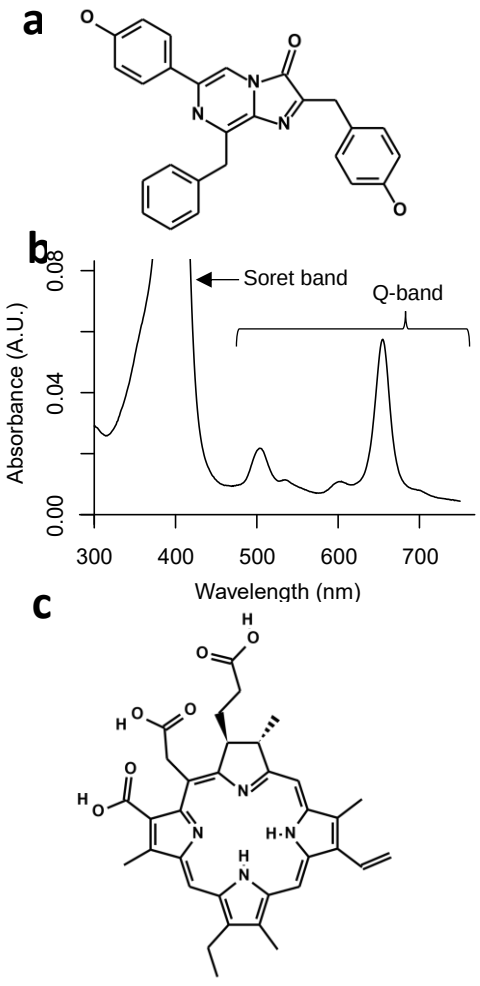


Figure 1.3: Compounds used by deep-sea fish. a) Structure of coelenterazine. **b)** Spectrum of chlorin e6, a porphyrin. **c)** Structure of chlorin e6.

The bioluminescence system dominant in marine ecosystems (Rees et al., 1998) is based on luciferases providing the right conditions for the chemiluminescence of coelenterazine (Figure 1.3a), a luciferin with an imidazopyrazine skeleton (Shimomura et al., 1980). It has maximum emission around 465 nm (Shimomura and Teranishi, 2000), and its emission profile is broader than the downwelling light. While there are adaptations that can be made to more finely tune the spectral profile produced, the original emission is very resistant to modification (Rees et al., 1998). Predators may have countermeasures exploiting this mismatch, such as pigmented lenses (Douglas and Thorpe, 1992) that are proposed to increase contrast between space light and mismatched camouflage, with gaps in the lens (Douglas et al., 1998b) or accessory retinae (Pointer et al., 2007) to compensate for loss of sensitivity when avoiding predators. Other adaptations include tapeta, which reflect light passing through the retina, giving photoreceptors a second chance to capture the light. Interestingly, an eye with a tapetum may not use it to simply boost signal, but increase signal to noise by making a concomitant reduction in receptor pigment density (Denton and Nicol, 1964), reducing thermal noise. Tapeta may also be pigmented, some in patterns adapted to the different sources of light (Shelton et al., 1992). A tapetum reduces the resolution of the

image due to scattering, but in an environment where light comes as approximately a point source (Warrant, 2000), resolution reducing-features such as a tapetum or a large pupil aperture carry less of a penalty.

Genus	Species	Method	Author	Year	Retinoid	λmax
Aristostomias	grimaldii	Retina extract	Bowmaker, Dartnall, Herring	1988	A1	517/518
			Bowmaker, Dartnall, Herring	1988	A2	552
	scintillans	Retina extract	Crescitelli	1991	A1	524
			O'Day, Fernandez	1974	A1	526
			O'Day, Fernandez	1974	A2	551
			Crescitelli	1991	A2	552
	tittmannii	Microspectrophotometry	Partridge, Douglas	1995	A1	518
			Partridge, Douglas	1995	A2	550
			Partridge, Douglas	1995	A1*	581
		Retina extract	Partridge, Douglas	1995	A1	523
			Douglas, Partridge, Marshall	1998	A1	523
			Partridge, Douglas	1995	A2	551
			Douglas, Partridge, Marshall	1998	A2	551
		Whole mount (fresh)	Partridge, Douglas	1995	A1	531
			Partridge, Douglas	1995	A1	531
			Partridge, Douglas	1995	A2	546
			Partridge, Douglas	1995	A2	550
			Partridge, Douglas	1995	A1*	586
			Partridge, Douglas	1995	A1*	590
		Whole mount (frozen)	Partridge, Douglas	1995	A1	527
			Partridge, Douglas	1995	A2	546
	xenostoma	Retina extract	Knowles and Dartnell	1977	A1	514
			Knowles and Dartnell	1977	A2	551
Malacosteus	danae	Retina extract	Crescitelli	1989	A1	514
			Crescitelli	1989	A2	556
	niger	Microspectrophotometry	Partridge et al.	1989	A1	521
		†	Bowmaker, Dartnall, Herring	1988	A1	522 (1.9)
		†	Bowmaker, Dartnall, Herring	1988	A1	522 (6.8)
			Partridge et al.	1989	A2	538
		†	Bowmaker, Dartnall, Herring	1988	A2	548 (3.6)
		†	Bowmaker, Dartnall, Herring	1988	A2	549 (5.5)
		Retina extract	Somiya	1982	A1	510
			Douglas et al	1998	A1	515
			Bowmaker, Dartnall, Herring	1988	A1	517/522
			Douglas et al	1998	A2	540
			Bowmaker, Dartnall, Herring	1988	A2	545
		Whole mount	Unclear		A1	522
			Unclear		A2	534
Pachystomias	microdon	Intact retina	Denton et al	1970	A2	575
		Microspectrophotometry	Partridge et al.	1989	A1	513
			Bowmaker, Dartnall, Herring	1988	A1	514
		†	Bowmaker, Dartnall, Herring	1988	A1	515
			Partridge et al.	1989	A2	539
			Bowmaker, Dartnall, Herring	1988	A2	541
		†	Bowmaker, Dartnall, Herring	1988	A2	544
		Retina extract	Partridge et al.	1989	A1	513
			Douglas, Partridge, Marshall	1998	A1	513
			Bowmaker, Dartnall, Herring	1988	A1	524/525
			Partridge et al.	1989	A2	539
			Bowmaker, Dartnall, Herring	1988	A2	539/540
			Douglas, Partridge, Marshall	1998	A2	540
		Whole mount(fresh)	Douglas, Partridge, Marshall	1998	A1	520
			Douglas, Partridge, Marshall	1998	A2	563
			Douglas, Partridge, Marshall	1998	A1*	595

Table 1: Comparison of opsins from dragonfish. λ_{max} coloured by wavelength: <530 nm : blue, 531 to 541 nm: green, 542 to 563 nm: orange, >564 nm: red. *indicates non-rhodopsin, †indicates difference microspectrophotometry

There are other uses for bioluminescence than camouflage. Deeper in the ocean, it becomes the dominant source of light, and a major method of all kinds of communication. It can be

used to signal between members of the same species (Herring, 2007), or in the same way bright colouring is used to signal danger (Grober, 1988) on land. Bioluminescence can be used to hunt, famously as a lure by the female anglerfish, or as a searchlight. Any exploitation of light can also be a vulnerability: if you can see it, so might the bigger fish. One interesting behaviour is that one fish may use other organisms' tapeta against them by using bioluminescence to look for eyeshine (Howland et al., 1992). Most organisms do, however, have a limited spectral sensitivity. Though the blue-green space light may not reach the depths, visual pigments mostly still use the same wavelengths. This is not to capture any possible light from the surface, but because the maximum transmission of light is at these wavelengths.

Dragonfish

The stomiidae, or barbeled dragonfish, are a family of fish in the mesopelagic zone, some of which have attracted attention for unusual vision (Bowmaker et al., 1988) and behaviour. One way some dragonfish avoid detection is to change wavelengths. In addition to normal, blue luminescence they have a second set of photophores which contain a fluorescent protein which changes the light to red, around 700 nm (Campbell and Herring, 1987; Widder et al., 1984). These have red shifted visual pigments (O'Day and Fernandez, 1974) compared to other fish (see table 1), as well as pigmented lenses and tapeta in some cases (Douglas et al., 1998b; Somiya, 1982). The red photophores have been proposed to work as a sort of invisible flashlight, allowing the dragonfish to hunt unseen. Though the propagation of this light would be over very short distances, due to the high attenuation of red light, even in the purest seawater (Smith and Baker, 1981), many deep sea organisms camouflage themselves by red colouring (Johnsen, 2005), proof against blue searches, but fodder for the dragonfish.

One of the dragonfish exploiting this strategy, *Malacosteus niger (M.niger),* also has a brown filter on its photophores (Widder et al., 1984) rendering its bioluminescence even dimmer in shortwave regions. It also may not have as far red-shifted visual pigments as other dragonfish. It was however, found to have deposits of a chlorophyll derivative (Douglas et al., 2016, 1999, 1998a) in its retina, which has been suggested ("Dragon fish see using chlorophyll | Nature") to work as an antenna for its visual pigment.

1.1.1.3: Chlorin

1.1.1.3.1: Porphyrins

The chlorophyll derivative found in *M. niger* eyes is one of the porphyrin family of pigments (Figure 1.3c). These compounds comprise a macrocycle of four rings modified by functional groups. There are two major groups of biological pigments, the carotenoids like retinal, and porphyrins. Their biological roles of porphyrins exploit their strong absorbance in the visible range due to the large conjugated system of the macrocycle, such as in chlorophyll; and their ability to coordinate metals using the nitrogens in the centre, such as in haem and related compounds. An unmodified porphyrin has two significant sets of absorbances, the B- or Soret band, which is found around 400 nm; and the Q-band, which is found between 480 and 700 nm (Figure 1.3b). Chlorins, which are porphyrins and include chlorophyll, can have additional strong absorbances in the far red. This is discussed in section 3.1. The far-red absorbance has been proposed to overlap with the emission from Malacosteid red photophores, then transferring the excitation to rhodopsin. This would require some form of uphill energy transfer, which has been observed in bacterial and insect systems (Balashov et al., 2005; Kirschfeld and Vogt, 1986). Possible mechanisms are discussed in section 1.1.1.3.2.

A further question is how a fish would acquire such a compound. Though the metabolisms of all animals include a porphyrin pathway used to produce haem, the pathways for haem and chlorophyll divide at protoporphyrin IX (Kegg pathway ko00860 (Kanehisa and Goto, 2000)). *M. niger*'s diet includes copepods (Sutton, 2005), zooplankton which commonly eat phytoplankton supposed to be the original source of chlorophyll digested to produce antenna compounds.

1.1.1.3.2: Experiments with Chlorin

Chlorins are used in photodynamic therapy (PDT) for cancer because of their high absorbance of infrared radiation (Zenkevich et al., 1996). There are anecdotal reports of PDT causing photosensitivity in humans ("Photodynamic Therapy for Cancer," n.d.), i.e. patients have reported increased sensitivity to light after treatment. This prompted investigation into the possibility that sensitization by porphyrins might not require a specific adaptation of the rhodopsin. The first such experiments were using bovine rhodopsin (Ilyas Washington et al., 2004), and showed a 2.9-fold enhancement in sensitivity to a narrow bandwidth of far red light. The same groups also administered the most promising compound, chlorin e6 (Ce6), to mice (Washington et al., 2007) and demonstrated both more bleaching by red light, and a larger electrical response from the retina; and have soaked salamander retinae in a Ce6

containing solution and demonstrated that rod bleaching by red light increased 180-fold (Isayama et al., 2006). A more questionable claim from this research is that Ce6 specifically targets the retina. While more Ce6 was found in the rods, specifically in their outer segments, which are very membrane dense, Ce6 is more soluble in lipids, which could explain this observation without resorting to targeting (Douglas et al., 2016). The observation of effects across species suggests that some form of enhancement of far-red sensitivity is, to some extent, independent of the rhodopsin.

If absorbance of red light by Ce6 does lead to activation of rhodopsin, some form of energy transfer should take place between Ce6 and rhodopsin. The absorbance of a photon by a system excites an electron from the ground state. Normally, the excited electron returns to the ground state by a combination of thermal relaxation processes. However, it is possible for some systems to thermally relax to the ground vibrational state of the excited state, then return to the ground state by emission of a photon of lower energy (this change, called the "Stokes shift", is due to the loss of energy by vibration). This process is fluorescence. It is also possible for the excited electron to have changed spin, which means it cannot directly return to the ground state by fluorescence. It has entered a forbidden "triplet" state. If it returns to ground state through another path involving emission of a photon, this emission is phosphorescence. There are processes by which the emitted energy from a donor can be transferred to an acceptor system. This can occur over long distances through resonance between the two in Förster resonance energy transfer (FRET) . The efficiency of the process depends on the sixth power of the separation between donor and acceptor, so the effect is exploited in microscopy and imaging applications (Jares-Erijman and Jovin, 2003). When FRET occurs, an alternative pathway for the return of the electron to the ground state is provided, so the measured fluorescence is decreased. Another mechanism by which energy transfer can take place is called Dexter energy transfer (DET) (Dexter, 1953). In this process, the excited electron itself is exchanged for a ground state electron, leaving a ground state donor and an excited acceptor.

Both FRET and DET normally require that there is a significant overlap between the emission of the donor and the absorbance of the acceptor. This is because the promotion of an electron to an excited state requires energy, which is proportional to the wavelength of absorbance by Planck's equation: $E = hv$, in which E is energy, h is Planck's constant, and v is the wavelength of a photon. From this relationship, the energy of the emitted photon must be close to the energy required to promote the electron, thus the wavelength of the emitted photon must be close to the wavelength of absorbance. As the donor loses energy through

relaxation, the energy transfer is also usually “downhill”, so the absorbance of the acceptor is at a longer wavelength than the absorbance of the donor. There are mechanisms which allow violation of this principle, such as the absorbance of multiple photons (Singh-Rachford and Castellano, 2010).

The normal requirement for downhill energy transfer would seem to forbid the transfer of the signal of a far-red photon from an absorbing porphyrin to isomerize the retinal of a rhodopsin with a λ_{max} near 500 nm. Previous experiments with rhodopsin and Ce6 also show that energy transfer cannot occur as an alternative to fluorescence emission, as the fluorescence of Ce6 is unaffected by the addition of rhodopsin (Washington et al., 2007). There remain mechanisms by which the energy transfer could still happen. DET processes can involve the transfer of triplet-state electrons from donor to acceptor. Ce6 has long-lived triplet states which are more similar in energy to triplet states of retinal than the singlet states are to one another. The efficiency of energy transfer between these triplet states would, therefore, be more efficient. It has also been observed that isomerisation of retinal can proceed via triplet states (Rosenfeld et al., 1974), so this remains as a possible mechanism for energy transfer.

Ce6 is not, however, simply an antenna pigment to rhodopsin which simply needs to be close enough to transfer excitation to retinal. Thermal stability of rhodopsin is improved by the addition of Ce6, and the effect increases with the addition of divalent metal ions, especially Zn^{2+} (F. Balem et al., 2009). Porphyrin metal binding is at the core of the function of haemoglobin and photosynthesis, so in some ways this is not surprising. Though in this paper the thermal stability measured was of the secondary structure of the opsin and not of the whole rhodopsin-retinal system, the fact that tapeta are used in fish to decrease dark noise and thereby enhance signal:noise suggests a similar possibility for chlorophyll derivatives.

It remains to be seen how relevant the different effects porphyrins exert on rhodopsin are in a physiological context, and whether any of these make it a useful therapeutic. Before any firm conclusions can be made, the red-bleaching enhancement effect must be systematically investigated. Previous attempts *in vitro* have made no effort to separate the contributions of Ce6 and rhodopsin in absorbance spectra. The behaviour of both species upon illumination with red-light should be characterised, and some method used to determine which effects in a mixture arise from which species. If a reproducible enhancement of bleaching by red light is found, some effort should be made to probe the nature of the interaction. There are fluorinated analogues of retinal (Colmenares, 1991) which provide a sensitive probe through

nuclear magnetic resonance (NMR, introduced in section 1.1.4.1). If the binding of Ce6 affects the retinal binding pocket, the NMR spectrum of fluorinated retinal should be affected.

Beyond these effects, the interaction of Ce6 with rhodopsin is of broader interest. The possible binding site for Ce6 includes the D/ERY motif (Woods et al., 2017), which is very highly conserved among rhodopsins and related proteins (Rovati et al., 2007). The conservation of this site between dragonfish opsin and the bovine opsin of the common *in vitro* experimental system should be explored, as a lack of conservation between the two could make the normal experimental system a poor analogue for the natural system. The proteins related to rhodopsin are physiologically and medically important, and, as the thermal stability of rhodopsin is increased by Ce6, the interaction is unlikely to be simply an energy transfer to retinal and is therefore potentially useful or informative for wider studies.

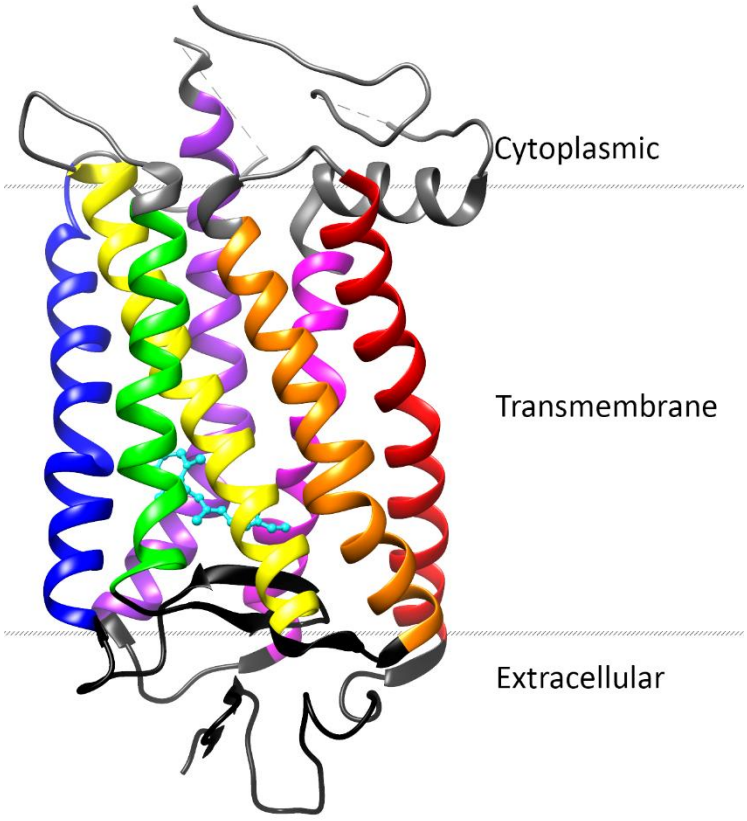


Figure 1.4: Structure of G protein-coupled receptors. The crystal structure of rhodopsin (PDB 1L9H) is used as a model. The extracellular domain is in black, the cytoplasmic in grey, and the transmembrane domain is coloured by helix. Retinal is represented as ball and sticks in cyan. Slanted dashed lines show the approximate ends of the membrane.

1.1.2: Rhodopsin as a folding model

Rhodopsin is the canonical member of the class A G protein coupled receptors (GPCRs). These respond to various stimuli in the same way rhodopsin responds to the conversion of 11-*cis*-retinal to all-*trans*, which is analogous to the binding of an agonist by a GPCR. Class A

GPCRs are targets for a third (Santos et al., 2017) of all pharmaceuticals and control many of the processes vital for an organism's health and survival. GPCRs are involved in responses to the outside world (e.g. rhodopsin for photons, odorant receptors (Malnic et al., 2004) for volatile chemicals) and signals generated in the body (e.g. chemokines (Krumm and Grisshammer, 2015), steroid hormones (Wang et al., 2015)). The activation of rhodopsin is relatively simple when compared to other GPCRs, whose activity can be heavily modulated, and exists on a spectrum of full and partial activation by different ligands (Hilger et al., 2018; Latorraca et al., 2017; Manglik and Kobilka, 2014).

GPCRs comprise seven α-helices spanning the membrane, connected by a series of loops which are essential to both their structure and function, ending in a short, eighth helix outside of the membrane and a flexible C-terminus. Another useful way of dividing a GPCR structure is to identify cytoplasmic, extracellular, and transmembrane domains (Figure 1.4).

Rhodopsin was the first GPCR to have its structure determined by x-ray crystallography (Palczewski et al., 2000). Crystal structures of other GPCRs have been determined since (Isberg et al., 2016), as have other states of rhodopsin (Ridge and Palczewski, 2007), which have been instrumental in understanding GPCR activation.

Rhodopsin is both unusually stable and incredibly sensitive, allowing the detection of single photons (Baylor et al., 1979). It has been successfully unfolded without aggregation *in vitro* by addition of 30% SDS (Dutta et al., 2010a). This was arrived at as the best denaturant for later studies by screening a series of denaturants and comparing extent of unfolding and the stability of the unfolded states (Dutta et al., 2010b). Rhodopsin in 30% SDS has lost 45% of its helical structure, as measured by circular dichroism, and remains soluble long term. These experiments are further discussed in section 5.2.1. Unfolded states are of interest as they relate to the folding and stability of proteins. As an unfolding protocol exists for rhodopsin, and the folded protein is so well-investigated and stable compared with its relatives, it makes an ideal model system for the folding of class A GPCRS.

There are diseases in which the unsuccessful folding of GPCRs is implicated, including reproductive disorders (Ulloa-Aguirre et al., 2014); some forms of Charcot-Marie-Tooth disease, a common neuropathy (Soong et al., 2013). The disease of rhodopsin misfolding is called retinitis pigmentosa.

1.1.3: Rhodopsin misfolding

1.1.3.1: Retinitis pigmentosa

Retinitis pigmentosa (RP) is a disease characterised by loss of night vision early in life, followed by loss of peripheral vision. The rod cells of the retina either fail to develop or die early. The death of the rods damages the retina, and later, even central vision can be lost. 15 to 35% of cases are inherited as an autosomal dominant condition (ADRP), and 20-30% of these are due to mutations in rhodopsin (Sandberg et al., 1995), of which more than 150 have been identified (Daiger et al., 2013; "RetNet: Disease Table"). Previous classifications of rhodopsin ADRP mutants have made divisions based on biochemical and cellular characteristics of mutants expressed in cell culture (Kaushal and Khorana, 1994), (Berson et al., 2002; Cideciyan et al., 1998; Sandberg et al., 1995), but more recent classification schemes have incorporated data from animal models and later established biochemical criteria to deliver seven classes of mutant (Athanasiou et al., 2018). These classes divide mutants by cellular dysfunction, stability, post-translational modification and disordered interactions with other proteins, so classes are not mutually exclusive.

Clinically, ADRP can also be classified according to clinical phenotype as either class A, with severe, early-onset loss of rod function; or class B, with a milder phenotype. These can be associated with different mutations, but there are other sources of variability (Iannaccone et al., 2006).

1.1.3.2: P23H

The P23H mutation is the most common ADRP mutation found in the United States, the cause of 10% of total cases of ADRP in north American populations of western European origin, and was the first point mutation to be associated with ADRP (Dryja et al., 1990). The clinical phenotype is class B, and within the biochemical and cellular classification system above is class 2, meaning it misfolds, is retained in the endoplasmic reticulum and is unstable, though this can be partially pharmacologically rescued by administration of 9-*cis* or 11-*cis*-retinal (Noorwez et al., 2004). P23H has been cloned into both transient and stable expression systems (Oprian et al., 1987; Reeves et al., 2002). Compared with wild-type rhodopsin, its yield of chromophore is low, though properly folded and misfolded P23H can be separated by control of elution conditions, and the folded fraction maintains WT-like secondary structure as measured by circular dichroism (CD) spectroscopy (Liu et al., 1996). However, P23H has compromised light response and thermal stability (Chen et al., 2014; Noorwez et al., 2004). As its yield can also be partially rescued by administration of 11-*cis* or

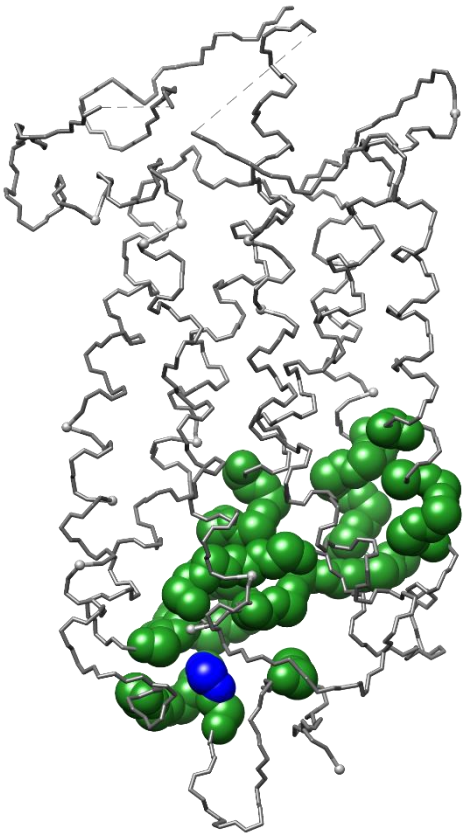


Figure 1.5: Folding core of rhodopsin. Residues of the folding core indicated as green spheres. P23 is shown as blue spheres

9-*cis* retinal during expression, it makes a good experimental system for rhodopsin misfolding. Any exploration of wild-type rhodopsin's unfolding states could be extended to P23H, providing an example of folding gone wrong. Knowledge of P23H misfolding would also be of clinical relevance due to the occurrence of retinitis pigmentosa.

The most widely used model for membrane protein folding does not account well for the folding behaviour of rhodopsin, but a newer model predicts a core of residues important for folding. The position 23 is found within the predicted folding core (Figure 1.5), and a proline at that position is well conserved within vertebrate and invertebrate opsins, as well as other GPCRs (Applebury and Hargrave, 1986; Dryja et al., 1990). The common model is the two-stage model, and the newer model, which accounts for aspects of rhodopsin's folding, is the long-range interactions model.

1.1.3.3: Two-stage model

The first membrane protein to be refolded from a fully unfolded state was bacteriorhodopsin (bR) from *Halobacteria salinarum* (Huang et al., 1981). This achievement made bR the principal folding model system for α-helical membrane proteins, especially 7-transmembrane helix proteins. That bR could also be refolded from fragments (Liao et al., 1983) to yield a protein indistinguishable from native bR (Popot et al., 1987) led in part to the formulation of the two-stage model for membrane protein folding (Popot and Engelman, 1990). This is the idea that folding of helical membrane proteins starts with the formation of independently stable helices, which then diffuse laterally in the membrane until they associate into their final arrangement. There are more fundamental reasons to suppose the formation of stable helices is independent of final protein folding. Helices form by the satisfaction of local hydrogen bonds. In a hydrophobic environment, hydrogen bonds cannot be satisfied by interactions with solvent, so such conditions make helices very stable. This model explicitly forbids the rearrangement of these helices post insertion, as they should be

so stable once inserted that this is impossible. There are proteins capable of the flipping of entire groups of helices (Vitrac et al., 2015; Woodall et al., 2015), and that have "frustrated" helices (Gafvelin and von Heijne, 1994) that are inserted or lie along the membrane under different conditions.

1.1.3.4: Long-range interactions model

Revisions to the two-stage model have been made (Engelman et al., 2003) to account for phenomena outside the partition and association of helices while maintaining these as the first and central events in helical membrane protein folding. While this paradigm is useful for understanding some proteins, there are others for which the observations cannot match up with the idea of independently stable helices. For these, there is a greater importance for long-range interactions; and loops, far from being irrelevances which can be cut apart with

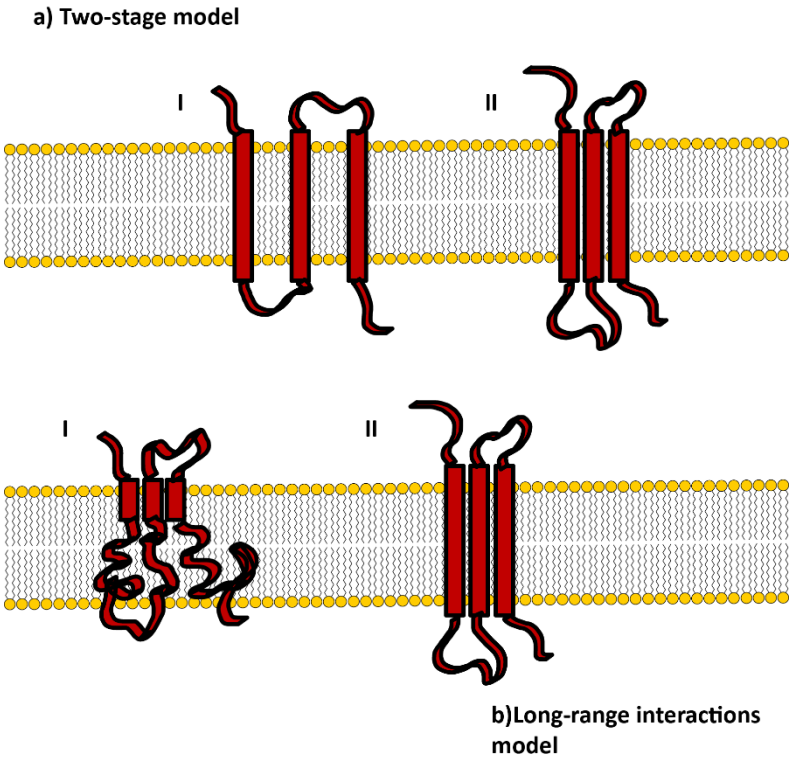


Figure 1.6: Two-stage model and the Long-range Interactions model. a) The two-stage model for membrane protein folding based on observation of bR. **I)** Independently stable helices form in the membrane. **II)** The helices condense to form a bundle. **b)** The long-range interactions model for membrane protein folding based on observation of MR. **I)** A folding core of loops and helix segments forms. **II)** The folding fore serves as a nucleus around which other structural elements form

impunity, are essential for proper folding. A protein which seems more compatible with a long-range interactions model for folding is mammalian rhodopsin (MR). A comparison between the models is shown in Figure 1.6.

The first evidence that MR might need a modified model for its folding mechanism came from co-expression studies of fragments which did not refold in the same simple fashion as bR (Ridge et al., 1995). More recently, the so-called Floppy inclusion and rigid substructure topography (FIRST) method of simulated unfolding was applied to MR and bR (Tastan et al., 2007). FIRST analysis uses theory from a long lineage of analysis of stability and of the flexibility of glasses to identify rigid regions of proteins (Jacobs et al., 2001; Rader et al., 2002).

The algorithm first uses the interatomic distances in a structure to model the bonds as a network of constraints on distances and angles. Regions which are "overconstrained", meaning that they have more bonds than are necessary to keep it rigid, some of which can be removed, are identified. Then "underconstrained" regions, which are rotatable parts between overconstrained regions, are identified. The energies of the hydrogen bonds are approximated and ranked by energy. They are then removed them in order of strength from the weakest (Hespenheide et al., 2002), which in the original application to glasses served to identify the point at which the glass melted. For proteins, this identifies a transition state in folding and finds folding cores, defined as the last rigid (or overconstrained) region containing two secondary structure elements.

This application of the FIRST algorithm was found to be in good agreement with experimental results using H-D exchange NMR spectroscopy to identify folding cores in soluble proteins (Rader and Bahar, 2004). This approach measures how quickly hydrogens in a protein can exchange with deuterium in the solvent. Regions which have slow exchange are more tightly constrained and form the folding cores of proteins (Li and Woodward, 1999). The same regions were identified when using the ranking function on the bond energies. If the ranking was not used and bonds were removed randomly, the results failed to align with experimental data, showing that the overall approach of identifying constraints on regions by bond energy is a valid approach for exploring protein folding.

When comparing bR and MR, it was found that FIRST fails to reliably identify a folding core in bR, which is in line with experimental results and the two-stage model (Tastan et al., 2007). It shows bR separating early on in simulated unfolding into its helices, which is appropriate for the two-stage model. MR, however, maintains a folding core between helices, suggesting that interactions between residues distant in sequence, but close in space in the folded protein, are important for its stability. The folding core also contains higher energy in the sum of its hydrogen bonds. These FIRST results provided a mechanistic explanation for why

the co-expression experiments with MR did not result in refolding of MR from fragments except for very specific cuts, and thus led to the proposal of a long-range interactions model (Klein-Seetharaman, 2005).

The folding core of rhodopsin is found in the extracellular domain (Figure 1.6), including residues in the extracellular loops and to the extracellular end of helices. Experimental data (discussed in section 5.2) shows that the extracellular domain remains more rigid than the rest of rhodopsin in unfolded states. Development of experiments which can show the dynamics of different domains in rhodopsin will test the predictions of a folding core.

1.1.4: Comparing rhodopsin and pSRII

When comparisons were made between bR and MR using FIRST analysis, the same analysis was carried out on other prokaryotic rhodopsins, including the sensory rhodopsin of *Natronamonas pharaonis* (Tastan et al., 2007). *Natronamonas pharaonis* is an archaebacterium. It produces a sensory rhodopsin (pSRII) which is used for phototaxis. Having high homology to bR, it can, like bR, act as a proton pump. Unlike bR, it also has a transducing partner. pSRII is the first 7-transmembrane helical membrane protein to have its structure determined by nuclear magnetic resonance (NMR) spectroscopy (Gautier et al., 2010), and its dynamics have been thoroughly explored by that technique.

1.1.4.1: NMR

Most protein structures are determined by X-ray crystallography (90% of structures deposited in the PDB were from x-ray crystallography at the time of writing). Though this is possible for membrane proteins, the need to incorporate some amphiphile to solubilise the molecule causes difficulty with crystallisation (Carpenter et al., 2008). Cryo-electron microscopy is another technique used for structure determination, but only recently has gained atomistic resolution (Kim and Sanbonmatsu, 2017). These techniques also only acquire limited information regarding the dynamics of molecules, as they often require low temperatures.

NMR, however, can be carried out at physiologically relevant temperatures, does not require crystallisation, and provides information about dynamics. These characteristics make it a useful technique for investigating membrane proteins (Opella and Marassi, 2017). A significant disadvantage is the requirement of large masses of sample. It is relatively insensitive, requiring milligrams of protein. Production of such amounts of membrane proteins is very challenging (Junge et al., 2008).

The signal detected by an NMR spectrometer arises from the spin of nuclei. As the nuclei are charged, their movement produces magnetism. When an external magnetic field is applied to an NMR sample, the magnetism of the nuclei aligns in the direction of the field. A magnetic pulse called the RF pulse is applied at right angles to the first, and the magnetization of the nuclei changes to that direction instead. As this pulse ends, the magnetization returns to align with the initial field. As it returns, it revolves around the direction of the magnetic field in a characteristic motion called "precession" at a frequency characteristic of that nucleus (the Larmor frequency). As the magnetization crosses a probe of the spectrometer, it induces a current, and as this occurs at the Larmor frequency, this can be detected by the instrument.

The magnetic field a nucleus actually experiences is not the applied magnetic field. Electrons, as moving charges, also respond to the applied field and influence the nuclei. Though this effect is small, it is detectable by NMR spectroscopy, and changes the resonant frequency of the nuclei. This phenomenon is known as chemical shift. Though these values have a value in Hertz (Hz), the values are small and depend on the strength of the applied field, so are reported as parts per million (ppm) relative to the frequency of a standard compound. As the environment the electrons produce is different for nuclei in all but symmetrical molecules, this allows NMR to discriminate between and identify nuclei.

Another effect that conveys information about the identity of peaks is that the spins of nuclei can be coupled. The spectra of coupled nuclei have peaks which are split into multiplets. This means that the peak is divided into multiple peaks which have the same area as a single peak for that nucleus. Multiplets of coupled nuclei are split by the same amount, which allows identification of which peaks belong to which nuclei. There are other through-bound and through-space interactions which can be used to assign peaks to nuclei and obtain spatial constraints between them.

Different ways of applying RF pulses can be used to detect different types of interaction between spins, so different pulse sequences are used in different NMR experiments which can be used alone or in combination to identify signals as belonging to particular nuclei. NMR experiments can be used to assign NMR peaks to residues in a protein, then connect the residues by bond (e.g. to ascertain primary structure) and through space (acquiring constraints on a protein's tertiary structure). Secondary structure can be found either through this through-space information or by characteristic changes in chemical shift (Wishart et al., 1992). Two experiments very commonly used in protein NMR are

heteronuclear single quantum correlation (HSQC) (Bodenhausen and Ruben, 1980) and heteronuclear multiple quantum correlation (HMQC) (Bax et al., 1983).

The "heteronuclear" part of each of their name refers to the fact that the experiment detects connected nuclei of different types, most commonly ^{1}H and either ^{13}C or ^{15}N. Both return a two-dimensional (2D) spectrum (Aue et al., 1976), with peaks that have both a ^{1}H and a heteronuclear component. These are commonly displayed as contour lines on a 2D graph. Both work by first exciting the ^{1}H nucleus, then transferring the magnetization to the heteronucleus, allowing the magnetization to evolve for a time, then transferring the magnetization back to the proton for detection. Though both experiments return roughly the same information, differences in the pulse sequences mean that HSQC spectra have higher resolution, while HMQC spectra can be easier to perform (a detailed comparison can be found in (Mandal and Majumdar, 2004). The two steps involving the ^{1}H nucleus deliver a stronger signal for the heteronucleus than excitation of that nucleus alone would produce, so these are relatively sensitive experiments.

While HSQC experiments generally deliver higher resolution than HMQC, some interactions within the methyl group of methionine make HMQC experiments up to three times more sensitive than HSQC (Tugarinov et al., 2003). Methyl groups in proteins are sensitive probes for structural biology of large proteins in any case, as they have three proton resonances in a very flexible group, resulting in slow relaxation (Ruschak and Kay, 2010).

Relaxation in NMR

The processes by which magnetization returns to the equilibrium, where all magnetization is along the axis of the applied field, are called relaxation. Faster relaxation results in a weaker, broader signal. The three protons and ε carbon of methionine relax slowly due to the group's flexibility and the three protons, which means they give a strong, sharp signal, which is very useful for proteins. Larger molecules in solution tumble more slowly, causing faster relaxation (Foster et al., 2007). This effect is exacerbated for membrane proteins, which require some membrane mimetic to stay soluble. The mass of the DM micelle solubilising rhodopsin is approximately 72 kDa (as measured by the manufacturer), increasing rhodopsin's effective mass from 32 kDa to 104 kDa.

Flexible parts of a protein relax more slowly than more rigid parts. There are methods which can precisely measure the different relaxation rates of specific sites in a protein, delivering details of the dynamics of the protein (Bracken, 2001; Ishima and Torchia, 2000). Less

precisely, the observation of narrower and broader peaks from normal experiments can be used to judge which sites are more or less dynamic.

Paramagnetic species in a sample will increase the rate of relaxation. Paramagnetism arises from the spin of unpaired electrons, which also align with the applied magnetic field. The dipolar interaction between the paramagnetic centre and a given nucleus is responsible for the increased rate of relaxation, and is dependent on the sixth power of the distances between the two. This means that signals from nuclei near to the paramagnetic species will weaken. The effect is normally limited to 35 Å, a much greater distance than other types of spatial information acquired from NMR.

A common source of paramagnetism for protein NMR experiments is the nitroxide radical, which contains an unpaired electron (Haugland et al., 2016; Sjodt and Clubb, 2017). There are reagents containing a nitroxide radical which will specifically react with surface cysteines of a protein. An example of the application of these labels to rhodopsin can be found in section 5.3.3.

1.1.4.2: Comparison of pSRII with bR and MR

pSRII takes an intermediate role between bR and MR: Based on its sequence and proton pumping activity, it is related to bR, but its transducing partner for signal transduction makes its role somewhat more like MR. Consistent with this intermediate role, FIRST analysis predicts that pSRII shares folding characteristics with both, as folding and conformational changes have been linked (Klein-Seetharaman, 2005; Tastan et al., 2014). Consistent with this comparison, FIRST analysis shows that the energy of the hydrogen bonds in its predicted folding core are somewhere between those found for MR and bR.

As bR and MR have already been established as model systems for membrane protein folding and display different folding mechanisms, it would be useful to establish another model system. pSRII shares the characteristics of binding retinal and comprising seven transmembrane helices with both and shares functional aspects with each. A comparison of pSRII with both will provide experimental evidence for the validity of applying FIRST analysis to membrane proteins and may be useful in the determination of which characteristics make a protein follow a two-stage or long-range interactions mechanism for folding. Whether its folding mechanism falls along some continuum between the two-stage folding of bR and the long-range interaction dependent folding of rhodopsin is a question with implications for the folding of membrane proteins in general. The retinal chromophore and previous NMR work on pSRII provide groundwork for the characterisation of pSRII as a model system.

1.2: Aims and objectives

1.2.2: Aims

The nature of the interaction between Ce6 and rhodopsin is unknown and could have implications for either the evolution of opsins, or the activation of GPCRs in general. The systematic replication of the observed enhancement of bleaching of rhodopsin by red light in the presence of Ce6 will be used to confirm an interaction between the two. The characterisation of this interaction will be attempted by measuring any effect Ce6 has upon retinal's chemical environment. MR and bR are already model systems for helical membrane protein folding but support different models for folding mechanisms. Finding the biophysical parameters for pSRII folding and establishing it as another model system with some characteristics in common with each may provide useful knowledge for the folding mechanisms of helical membrane proteins in general. MR will be further developed as a model system for membrane protein folding by characterising the NMR spectra of rhodopsin with isotope labelled methionines so that further experiments may explore the dynamics of different domains in unfolded states. Such new insight could have clinical relevance for the design of future approaches to treat retinitis pigmentosa.

1.2.3: Objectives

1. To probe the nature of the interaction between Ce6 and rhodopsin, bleaching experiments were conducted, recording the absorbance spectra of Ce6 and rhodopsin mixtures after irradiation with red light. The contributions of different concentrations of Ce6 and rhodopsin were separated during data analysis. The same data analysis approach was applied to different treatments of these mixtures to explore the effect of Ce6 on rhodopsin's stability. The conservation of residues in a 151-sequence multiple sequence alignment (MSA) was compared at different positions of interest. Samples of rhodopsin were reconstituted with 14-fluororetinal for ^{19}F NMR spectroscopy to see if Ce6 had any effect on retinal's environment.
2. To investigate the unfolding reactions of pSRII, in collaboration with the University of Cambridge, an experimental system unfolding pSRII in sodium dodecyl sulphate (SDS) was established based on previous work on MR. The nature of these unfolding reactions was explored by decomposition of absorbance spectra and fluorescence time courses to establish kinetic parameters, using non-linear least squares (NLS) fitting. This data analysis was expanded to analyse refolding

reactions and unfolding in the presence of hydroxylamine, which modifies the kinetics. More detailed decomposition methods were also used to examine the changes in pSRII absorbance spectra in non-denaturing concentrations of SDS.

3. Several data analysis techniques were applied to find underlying patterns in the changes of pSRII NMR spectra unfolded in different concentrations of SDS. Principal Components Analysis (PCA) was performed on a reduced dataset, and preliminary attempts were made to apply image analysis and clustering techniques.
4. To establish a system to further characterise the nature of unfolded states of rhodopsin in SDS, ^{13}C, ^{15}N-methionine labelled rhodopsin samples were prepared for NMR spectroscopy. HMQC spectra were recorded in various conditions in collaboration with the University of Cambridge and analysis of peak positions was used to assign peaks to rhodopsin methionines. Labelling with paramagnetic relaxation enhancement (PRE) reagents was attempted to assist ambiguous assignments.
5. Samples of P23H rhodopsin were prepared as for WT as above so comparisons might be made. Density filtering and peak fitting algorithms were used to determine the weight and estimate the density of Western blots scanned to saturation and compare the deglycosylation of wild-type rhodopsin with the P23H and N15S mutants.

CHAPTER 2: MATERIALS AND METHODS

2.1: Materials

- HEK293T cells from Dr R Murthy (University of Warwick)
- HEK293S with tetracycline inducible expression of rhodopsin
    - Expressing Wild-type a gift from Dr P Reeves, University of Essex (Reeves et al., 2002)
    - Expressing P23H from University of Pittsburgh (unpublished results)
- Rod outer segment (ROS) suspension
- Chlorin e6 (Ce6) from Frontier Scientific
- 4,4'-dipyridyl disulphide (PDS) from Sigma Aldrich
- S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) from Santa Cruz Biotechnology
- Retinal
    - 11-*cis* from Dr R Crouch (Medical University of South Carolina)
    - 9-*cis* from Sigma Aldrich
- TETSQVAPA nonapeptide from GenScript
- Sodium dodecyl sulphate (SDS) from Fisher Scientific
- n-dodecyl β-D-maltoside (DM) from Anatrace
- Lauryl Maltose Neopentyl Glycol (LMNG) from Generon
- Amphipol A8-35 from Anatrace
- Tandem amphiphile (AM-2) a gift from Dr A Marsh (University of Warwick)
- Foetal bovine serum (FBS) from Gibco
- Pluronic F-68 from Applichem
- Heparin (sodium salt) from Applichem
- Sodium butyrate from Alfa Aesar
- Glucose from Applichem for tissue culture, otherwise Fisher Scientific
- Cyanogen bromide-activated Sepharose 4B from Sigma Aldrich
- 4-chloro-1-naphthol from Sigma Aldrich
- Geneticin (G418) from Cellgro
- Hydroxylamine hydrochloride from Sigma Aldrich
- 2-(N-morpholino)ethanesulfonic acid (MES) from Sigma Aldrich
- PNGase F kit from New England Biolabs
- Color Prestained Protein Standard, Broad Range (11–245 kDa) from New England Biolabs

- 1D4 antibody from Cell essentials
- All other antibodies from Santa Cruz Biotechnology unless otherwise stated
- Competent *E. coli* (BL21) from New England Biolabs
- Vivaspin 20 concentrations, 30 kDa cutoff from Sigma Aldrich
- Visking tubing with 12 kDa cutoff, 21.5mm diameter from Medicell
- All other materials from Fisher Scientific unless otherwise stated

Buffers and media

- Sodium phosphate (NaPi) buffer (pH 6)
    - Ratio of 20:124 Na_2HPO_4:NaH_2PO_4
    - Adjust pH with phosphoric acid/sodium hydroxide
    - Used at 2mM unless otherwise stated
- Phosphate buffered saline (PBS)
    - 8 g/L NaCl
    - 0.2 g/L KCl
    - 1.44 g/L Na_2HPO_4
    - 0.24 g/L KH_2PO_4
    - Adjusted to pH 7.4 with HCl
    - Either taken from University of Warwick media prep (1X, autoclaved) or prepared as a 10X stock and diluted.
- From University of Warwick media preparation service
    - Dulbecco's Modified Eagle Media (DMEM)

      From Gibco, high glucose DMEM (catalogue number 41965), buffered with $NaHCO_3$
    - DMEM as above, but deficient in methionine. Buffered with 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)
    - 10000 I.U./mL / 10000 µg/mL Penicillin/Streptomycin (formedium)
    - 10 mg/ml Tetracycline (formedium)
    - Trypsin/EDTA in HBSS -Ca, -Mg
        - 8 g/L NaCl
        - 0.4 g/L KCl
        - 48 mg/L Na_2HPO_4
        - 1 g/L glucose
        - 60 mg/L KH_2PO_4
        - Phenol red to 1 part in 100,000

        - 2.5 g/L trypsin(1:250)
        - 0.38 g/L ethylaminediaminetetracetic acid (EDTA)
        - 0.35 g/L $NaHCO_3$
    - Super optimal medium with catabolite repression (SOC) medium as in (Hanahan, 1983)
- 10x running buffer for SDS polyacrylamide gel electrophoresis (SDS-PAGE) (250 mM Tris, 1.92 M glycine, 1% SDS, pH 8.3):
        - Tris base 30.30 g
        - Glycine 144.10 g
        - SDS 10.00 g
        - H2O to 1 L
    - Dilute 100 ml 10x buffer to 1 L with sterile water
- Gel for SDS-PAGE

Reagent	Resolving gel (per ml)	Stacking Gel (4%)(per ml)
30 % Bis/Acrylamide	X/30 ml	133 µl
Tris-HCl	250 µl	252 µl
10% SDS	10 µl	10 µl
H_2O	734.5 µl -X/30ml	599 µl
TEMED	0.5 µl	1 µl
10% APS	5 µl	5 µl

    For an X% gel
    - Tetramethylethylenediamine (TEMED)
    - Ammonium persulphate (APS)
- Western blot transfer buffer
    - 1X Tris-Glycine (30g/L Tris, 144g/L glycine)
    - 20% methanol
- Tris-buffered saline/Tween-20 (TBS-T)
    - 50mM Tris
    - 150mM NaCl
    - 0.1% Tween-20
    - Adjust pH to 7.4 with HCl
- Blocking Solution for western blots
    - 3% w/v non-fat milk

- 1X TBS-T

- Antibody solutions
  - Blocking solution (as above)
  - Required dilution of antibody
    - 1:2000 for 1D4
    - 1:2000 for anti-mouse

- Developing solution
  - 25ml PBS
  - 175 µl 3% 4-Chloro-1-Naphthol
  - 25 µl H_2O_2

2.2: Methods

2.2.1: Rhodopsin production

Rhodopsin samples were prepared from both rod outer segment suspension and from transfected cells. The HEK293T cell line is derived from human embryonic kidney cells, and expresses the SV40 large T antigen, allowing high levels of expression when transfected with plasmids containing the SV40 origin of replication (DuBridge et al., 1987). These were used for transient transfections, in which the cells are harvested shortly after transfection. The pMT4 plasmid used to express rhodopsin in HEK293T cells contains the SV40 origin of replication, first used to express rhodopsin in Oprian et al., 1987 using the p91023 plasmid (Kaufman, 1985; Wong et al., 1985). The pMT4 plasmid is modified from the pMT2 plasmid to contain the NotI restriction site (Franke et al., 1988) to make cloning of the opsin gene easier (the pMT2 plasmid has a PstI site in that place, which also exists within the opsin gene), which in turn was modified from the p91023 plasmid to confer β-lactam resistance instead of tetracycline resistance.

HEK293S cells are adapted for culture in solution, which allows higher density of cultured cells. HEK293S cells stably transfected with a tetracycline inducible expression system were used to produce samples in the milligram range (Reeves et al., 2002). It is possible to label protein produced from cell culture with amino acids containing isotopes useful for NMR spectroscopy, an approach used here to prepare NMR samples of rhodopsin.

2.2.1.1: Cell culture

2.2.1.1: Media

DMEM was used for all cell culture. To this, FBS was added to 10%. FBS contains growth factors and other components essential for culturing eukaryotic cells. When maintaining stably transfected cells, a maintenance concentration of 250 µg/mL G418 was added, which

provides a selection pressure for the cells to keep the plasmid. For cells in dishes, supplementation with antibiotics was not usually necessary and avoided if possible.

In preparation of NMR media

^{13}C, ^{15}N-methionine was added to methionine deficient media to 30 mg/mL from a 100X stock which had been filtered through a 0.22 µm filter. FBS used for this media was dialysed to remove free amino acids and small peptides to limit unlabelled methionine contamination of the media. Visking tubing with a molecular weight cutoff of 12 kDa was cut to ~20 cm lengths, washed then filled with distilled water and submerged in distilled water for 1h, then filled with ~30 mL FBS. The FBS was dialysed against 4x2 L PBS, changing at 30 minutes, 1h, 2h and 4h then left overnight. The dialysed FBS was filter sterilised.

2.2.1.2: Small scale culture of cells

HEK293T and HEK293S cells were maintained in 10 cm tissue culture dishes with 10 mL DMEM+10% FBS. Tissue culture took place in a class 2 laminar flow hood unless otherwise stated. Cells were incubated at 37°C and 5% CO_2. When judged by microscope-aided eye to be 80-90% confluent, they were passaged by first aspirating the media, then rinsing with 5 mL autoclaved PBS and adding 2 mL trypsin/EDTA, which was swirled around the plate to ensure good coverage. This was incubated at 37°C for 30 seconds, then examined under microscopes. If the cells had rounded up and detached, 8 mL DMEM was added, then they were taken up and down in a 10 mL disposable pipette until homogenous. This suspension was either decanted into plates as required and the medium supplemented, or used for spinner flask culture.

2.2.1.3: Large scale culture of cells in spinner flasks

For large quantities of rhodopsin, inducible cell lines were grown in 2 L spinner flasks (Corning). Confluent cells in 10 cm dishes were passaged and split into 15 cm dishes to a final dilution of 1:5 by volume, or 2:5 by plate as a 15 cm plate was given 20 mL media. Once the 15 cm dishes reached confluency, for each spinner flask, 3 were passaged and used to inoculate 500 mL media in the flask. Media for the flasks was supplemented with 0.1% pluronic F-68 (a surfactant which reduces damage from the shear forces introduced by spinners), 500 µg/mL heparin (sodium salt) (a polysaccharide which encourages cell growth) and 1:200 penicillin/streptomycin. Spinner flasks were incubated at 37°C on a stirrer plate set to approximately 20 revolutions per minute (rpm). The stirring of the spinners aerates the culture, ensuring oxygen delivery to the cells.

2.2.1.4: Freezing of cells for long term storage

When storage of the cell lines was required, confluent cells were passaged and when detached from the dish were centrifuged at 1000 rpm for 10 minutes in an Eppendorf 5810 centrifuge. The supernatant was aspirated and the pellet gently resuspended in 2 mL DMEM+10% dimethyl sulphoxide (DMSO). When DMSO is added to water, the ice crystals formed in freezing are smaller. This protects cells from bursting while freezing. This was divided into two cryo-tubes and placed in a -80^{o}C freezer in a polystyrene container overnight. For long term storage, these were subsequently moved into liquid nitrogen.

To revive frozen cells, the cryo-tubes were heated in a 37^{o}C waterbath for about 60 seconds or until a small lump of ice remained. If the cells were taken from liquid nitrogen, the tube was first opened in the laminar flow cabinet to release any liquid nitrogen which could explode the tube. The contents of the cryo-tube were transferred into a 15 mL falcon tube and diluted with 5 mL media dropwise. This was centrifuged at 1000 rpm for 10 minutes in an Eppendorf 5810R centrifuge. The supernatant was aspirated and the pellet resuspended in 5 mL media. This was added to 10 mL media in a 10 cm dish and swirled gently. The media was changed once the cells had attached (usually 4 to 6 hours later) to minimise DMSO exposure.

2.2.1.6: Calcium phosphate transfection

When small quantities of rhodopsin were required, HEK293T cells were transiently transfected with pMT4 plasmids containing rhodopsin. This was achieved using the calcium phosphate method, which forms precipitates between the phosphates of the DNA backbone and calcium introduced in the protocol. The precipitates settle against the membranes of cells and are endocytosed. The method is reliable, but requires a large amount of DNA, so is only appropriate for small quantities. 24 hours after passage, media was changed. Values given here are for one plate. To transfect multiple plates weights and volumes are simply multiplied by the number of plates. DNA solution equivalent to 20 µg was made up to 450 µl with sterile water. 50 µl 2.5M $CaCl_2$ was added dropwise, shaking between drops and shaking to mix thoroughly once all $CaCl_2$ was added. 500 µl HEPES buffered saline was added dropwise. The mixture was left to stand at room temperature for 1 minute. 1 ml of transfection mixture was added dropwise evenly across a plate, which was then swirled to mix. Plates were incubated for 48 hours before harvesting.

2.2.1.7: Induction of stable lines for NMR samples of rhodopsin

Stable lines used were inducible by addition of 2 µg/mL tetracycline. Spinner flasks grown for 5 or 6 days until media had turned orange were supplemented with 100 µL 10 mg/mL

tetracycline and 5 mL 500 mM sodium butyrate. Sodium butyrate has several effects upon cells which are useful for protein production in cell culture. Primarily, it inhibits histone acetylation, increasing the expression of some genes, and increases yield from these cell lines (Davie, 2003; Kruh, 1981). It is essential for efficient production of rhodopsin using the tetracycline inducible system (Reeves et al., 2002) They were also fed each day post induction with 6 mL 20% (w/v) glucose and 8% (w/v) sodium bicarbonate was added for pH balancing, estimated by the colour of phenol red in the media. Cells expressing P23H were also supplemented at 2 and 24 hours with 50 µL 100 mM 9-*cis* retinal to improve expression (Noorwez et al., 2004; Saliba et al., 2002). It was also found that without retinal supplementation, induced cells in a dish would die (data not shown). Cells supplemented with retinal were wrapped in foil, so pH correction was not possible. Cells were grown for 48 hours before harvesting.

2.2.1.8: Harvesting of Rhodopsin

2.2.1.8.1: Preparation of samples in dodecyl maltoside

Dishes were first washed with 10 mL ice-cold PBS, then 2 mL PBS was added and the cells were removed with scrapers. A 5 mL pipette was used to resuspend the cells, which were added to a falcon tube on ice. Another 2 mL PBS was used to wash each plate into the falcon tube. Tubes were then centrifuged at 4000 rpm for 10 minutes at 4^{o}C in an Eppendorf centrifuge with rotor. Supernatant was aspirated and the cell pellet was resuspended in 0.9 mL PBS per plate. 0.5 µL retinal per plate harvested was added to cell suspension from a 100 mM stock in absolute ethanol. The tubes were wrapped in foil and agitated for 90 minutes. DM was added to cell suspension to 1% from a 10% (w/v) stock. For dishes, DM was added to the whole solution after one addition of retinal.

Cell suspension from spinner flasks was poured directly into 500 mL centrifuge bottles. Each flask was rinsed with 2 x 50 mL ice-cold PBS, which was added to the centrifuge bottles. These were centrifuged at 4000 rpm for 10 minutes at 4^{o}C in a fiberlite F10BCI-6X500Y rotor. The supernatant was discarded and cell pellets resuspended in 25 mL ice-cold PBS and decanted into 50 mL falcon tubes. The bottles were washed with 25 mL ice-cold PBS. This was then centrifuged at 4000 rpm for 10 minutes at 4^{o}C in an Eppendorf centrifuge. The supernatant was discarded and the cells were resuspended in 18 mL ice-cold PBS per flask. In solubilisation, each flask was treated as 20 dishes. However, due to the large potential yields of rhodopsin from the stable cell lines, after retinal had been added and the sample agitated for 90 minutes, a 1 mL aliquot was taken and DM added, and retinal was added to the sample.

This was repeated until no further increase in rhodopsin was observed by absorbance spectroscopy, then DM was added to the whole sample.

Once reconstitution and solubilisation were complete, samples were centrifuged at around 67000 g. For small samples, this was at 35000 rpm in a Beckman TLA100.3 rotor; for larger samples, a Beckman 60Ti rotor was used at 25000 rpm.

2.2.1.8.2: Solubilisation of rhodopsin for LMNG samples

DM creates a large background signal in NMR which conceals some of the peaks of rhodopsin. LMNG creates less of a background, so was used to solubilize rhodopsin for some NMR experiments. The efficiency of LMNG solubilisation of rhodopsin was tested using ROS suspension as described below. Yields were found to be lower (as shown in section 5.3.1) , so LMNG samples were solubilised as above.

2.2.1.2: Harvesting of rhodopsin from bovine retinae

10% (w/v) DM stock was added to ROS suspension to 1%, and nutated end over end for 90 minutes. A difference spectrum (see below) was taken to determine the concentration of rhodopsin in the sample and purified as described below.

When testing the ability of different amphiphiles to solubilise rhodopsin, two 900µL aliquots of ROS were used for each. 100 µL 10% (w/v) stock solution of amphiphile was added to each and the aliquots were nutated end over end for 90 minutes. The samples were centrifuged at 15000 rpm in a Beckman TLA 100.3 rotor, and the supernatant collected. The pellet was resuspended in PBS+1% DM and nutated end over end for 90 minutes. This was then pelleted at 35000 rpm in the same rotor and the supernatant collected. The rhodopsin content of the original supernatant and the pellet's supernatant were measured by difference spectroscopy as described in section 2.2.2.2.

2.2.1.3: Preparation of 1D4-sepharose beads for rhodopsin purification

Rhodopsin was purified by affinity chromatography to the 1D4 monoclonal antibody raised against the C-terminal nonapeptide of rhodopsin, TETSQVAPA. The antibody was coupled to sepharose by the cyanogen bromide method. Cyanogen bromide activated beads were weighed out as a powder. Each gram of powder would bind 25-50 mg antibody, and have a final volume of 3.5 mL as a slurry.

The powder was suspended in 1 mM HCl, then washed in 200 mL 1 mM HCl per gram on filter paper in a Buchner funnel. The HCl was added in several aliquots, drawn through carefully with a vacuum pump and not allowed to dry completely. The antibody was supplied in 0.1 M $NaHCO_3$ + 0.5 M NaCl, pH 8.3, and was at a concentration of 3 mg/mL. The ideal concentration

of antibody for binding is between 5 and 10 mg/mL. As the antibody concentration was 3 mg/mL, it was decided that concentrating the antibody would have been inefficient, so the antibody was used as supplied. The desired mass of antibody for the mass of powder was added to the prepared beads in a stoppered flask or a falcon tube and nutated end over end. Each hour, the mixture was centrifuged and an aliquot of the bead-free supernatant was taken. The A_{280} of this aliquot was measured to assess the unbound protein concentration using the extinction coefficient of 1.4 Absorbance = 1.0 mg 1D4. Once 96% of antibody had been bound, excess ligand was washed away on filter paper on the Buchner funnel using at least five bead volumes of coupling buffer. Remaining CNBr groups were blocked by transferring the beads to 3.5 mL per gram powder 0.1 M Tris-HCl, pH 8, left to stand overnight at room temperature.

The beads were then washed in 3 cycles of 5 volumes, alternating between 0.1 M acetic acid/sodium acetate + 0.5 M NaCl at pH 4 and 0.1 M Tris-HCl + 0.5 M NaCl at pH 8. The beads were then transferred to PBS as a 50% slurry.

The binding capacity of the beads was assessed by taking a sample of rhodopsin of a known concentration, then adding enough beads to bind half of that rhodopsin, assuming a bead capacity of 1 mg rhodopsin per mL beads. The purification protocol was followed, and the fraction of rhodopsin bound measured. If all the rhodopsin in the sample was bound, a smaller volume of beads was added and the measurements repeated.

2.2.1.4: Purification of rhodopsin

2.2.1.4.1: Binding to beads

Once the rhodopsin concentration of a sample had been determined by absorbance spectroscopy, the volume of bead slurry was added to bind the amount of rhodopsin in the sample +10% to ensure all rhodopsin was bound. This was agitated overnight or for more than six hours at 4^{o}C. After rhodopsin binding to the beads, the mixture was added to a column. For bed volumes over 1 mL, a reusable column was used, and for smaller bed volumes, a disposable 10 mL column was used

2.2.1.4.2: Elution

The column was washed with 50 column volumes of PBS + 0.05% DM, then, if using NaPi to elute, 10 bead volumes of NaPi+0.05% DM. For 10 mL columns, the flow was not controlled, and for reusable columns, the flow rate was set between 0.5-1 mL/min. Protein was eluted with NaPi + 0.05% DM with 70 µM nonapeptide. For 10 mL columns, five initial elutions of 500 µL were made. The first elution was done without first incubating, and all subsequent

elutions were done after 20 minutes of incubation. For reusable columns, seven initial elutions of 4 mL were made with a flow rate of 0.2-0.5 mL/min. Absorbance spectra were taken of each fraction and the A_{500} used to calculate concentration. If the A_{500} of the last elution was over 0.02, further elutions were made.

2.2.1.4.2.1: Elution of P23H

Elution of P23H rhodopsin was carried out as above, but NaPi was found to be an ineffective buffer for elution, as only small yields of protein were eluted when using NaPi. As a higher salt concentration will elute more rhodopsin, NaPi was replaced with PBS. Though this increases the elution of misfolded rhodopsin, yields of P23H were prohibitively low when using NaPi.

2.2.1.4.2.2: Elution with LMNG

Rhodopsin samples in LMNG were first solubilised in DM, then eluted in PBS + 0.01% LMNG. NaPi + 0.001% LMNG was found to be ineffective, yielding a low proportion of the rhodopsin to which beads had been added.

2.2.1.5: Concentration of samples for NMR spectroscopy

To raise the concentration of eluted samples for NMR spectroscopy, they were concentrated using centrifugal filtration units with a cutoff of 30 kDa. Eluted fractions were added 4 mL at a time and centrifuged at 4000 rpm, in the dark, initially for 10 minutes. As the concentration rose, the concentration of detergent also rose, and the time taken for a given volume of solution to clear the column increased, so centrifugation time was increased as necessary. Once all the desired rhodopsin was in the filtration unit, the sample was washed with 4x1 mL NMR buffer. The sample was resuspended in the final desired volume of NMR buffer and transferred to NMR sample tubes. Concentration of rhodopsin and of detergent was calculated from their respective quantities in the samples added and the final volume achieved.

2.2.1.6: Production of DNA for transfections

Top10 *E. coli* were transformed with pMT4 plasmids containing rhodopsin by heat shock. 1-2 x 10^9 cells were mixed with 100 ng DNA in 50 mM Tris-HCl (pH 8). These were placed in a 42^{o}C water bath for 45 seconds, then cooled on ice for 2 minutes. The mixture was made up to 1 mL with SOC media and placed in a shaking incubator at 37^{o}C for 1 hour, then 200 µL was spread onto 10 cm plates of LBA ampicillin and incubated at 37^{o}C overnight. Colonies were picked and used to inoculate 500 mL LB medium with 100 µg/L ampicillin. This was incubated overnight, then used with a Maxiprep kit (Qiagen) to the manufacturer's instructions.

2.2.2: Absorbance spectroscopy

2.2.2.1: Data collection

The retinal chromophore of both rhodopsin and pSRII absorbs light strongly at ~498nm. This is useful for biophysical characterisation of these proteins, as the sample requirements are relatively low (rhodopsin samples of 20 µg/mL and below are usable for absorbance experiments), and the measurements are quick and relatively simple to perform. The ratio of the 500 nm peak and the 280 nm peak from the absorbance of aromatic residues gives an idea of the quality of purified samples, as only properly folded rhodopsin has 500 nm absorbance. Activation of rhodopsin or denaturation of rhodopsin or pSRII causes a loss of 500 nm absorbance and a gain in 380 nm absorbance. Though absorbance experiments alone cannot deliver precise information about these proteins, the experiments provide some insight into the proteins' tertiary structure.

Absorbance spectra were recorded on a Perkin-Elmer Lambda 35 spectrophotometer. Most spectra were recorded with a scan speed of 480 nm/min, though time courses with fast cycle times were recorded with a scan speed of 960 nm/min. Unless specified, all spectra were recorded from 650 to 200 nm. Measurements requiring temperature control were made using a water jacketed cell holder and a circulating waterbath. For experiments requiring side-by-side comparison of time courses, the water jacketed cell holder is capable of changing cells without opening the sample chamber.

Absorbance spectra of the pSRII samples were recorded at the University of Cambridge using a UV-1800 spectrophotometer (Shimadzu). Spectra were recorded between 250 and 600 nm with a fast scan rate (accumulation time = 0.05 s) and a sampling interval of 1 nm.

2.2.2.2: Difference spectra

For samples with high background such as cell lysates, difference spectra were recorded. To this effect, a spectrum was taken, the cuvette was illuminated with a Schott KL 1500 cold light source with an open aperture and colour temperature set to 3400 K for 60 seconds, using a fibre optic accessory. A second spectrum was recorded and the difference in A_{500} used to measure rhodopsin concentration.

2.2.3: Biochemical assays

2.2.3.1: Experiments with Ce6

2.2.3.1.1: Bleaching experiments

The interaction between Ce6 and rhodopsin, it is claimed, causes an increase in the bleaching of rhodopsin at 500 nm when illuminated with red light. To test this, samples of rhodopsin

mixed with different concentrations of Ce6 were illuminated with light shone through a longpass filter (a filter which allows light with a longer wavelength than its cutoff to pass through). The rates of bleaching at 500 nm after the subtraction of the Ce6 in the spectrum were used to establish whether an increase could be observed.

Rhodopsin in 2 mM sodium phosphate pH6, 0.05% DM was diluted to 400 µL at ~1 µM, and a spectrum recorded from 750 to 250 nm. Ce6 was added from 100 mM stock, or a dilution thereof, in DMSO so the volume added was between 1 and 4 µL, at the final concentrations indicated. The same protocol was followed adding 4 µL DMSO alone to obtain a baseline. An initial spectrum was recorded. The cuvette was illuminated with a Schott KL 1500 light source with an open aperture set to 2900 K through a fibre-optic accessory and a 630 nm long-pass filter (HV Skan) for 30 seconds, then another spectrum recorded. This was repeated until the sample had been illuminated for five minutes. For two samples, to confirm no bleaching was due to ambient light, the same procedure was followed, but covering the end of the fibre-optic, and no bleaching occurred.

After five minutes of red illumination, the 630 nm filter was exchanged for a 495 nm filter and the light source was set to 3400 K, and the sample illuminated for 60 seconds for full bleaching of rhodopsin.

Ce6 was added to the same buffer, omitting rhodopsin, from a DMSO stock as above. Such samples were diluted with buffer, illuminated as above, or had SDS added to 3 and 0.5% (w/v). The sample in 3% SDS was also diluted with buffer to test the effects of reducing SDS concentration.

Samples either diluted or illuminated were fitted as a multiplication of the original spectrum (details in section 2.3 below). The diluted spectra were recorded to ensure that the fitting protocol was appropriate to test the linear response of recorded spectra to reduced Ce6 concentration. These spectra, after multiplication by the fitted value, matched the original spectra very precisely (data in Chapter 3), so residuals fitting an illuminated spectrum to the original arise from uneven bleaching as discussed in results. The same procedure was followed when diluting SDS treated samples with buffer, and assumed to be valid for the same reasons.

2.2.3.1.2: Dark noise experiments to test stability of rhodopsin with Ce6

To ensure that only the light shone through the longpass filter was causing bleaching of rhodopsin when mixed with Ce6, rhodopsin was diluted and Ce6 added as for bleaching experiments. The cuvettes were in a water jacketed cell holder with the water set to the

desired temperature. Spectra were recorded every 20 minutes for 420 minutes. A final difference spectrum was obtained to determine the exact concentration of rhodopsin. Remaining rhodopsin over time was plotted as a fraction of the difference between the first spectrum and the illuminated spectrum.

2.2.3.1.3: Reconstitution of rhodopsin with 14-fluororetinal

Rhodopsin was reconstituted with 9-*cis*-14-fluororetinal to introduce a ^{19}F label in retinal for NMR studies. The reconstitution protocol was designed to allow for the reaction to take place in an NMR sample tube so it could be followed by NMR spectroscopy. A rhodopsin spectrum was measured, the cuvette was illuminated with the Schott KL 1500 cold light source as described, and a second spectrum was taken. Fluororetinal was initially added so that the absorbance at 380 nm was roughly equal to the difference in A_{500} brought about by illumination, though, later, multiples of this value were added. It was added from a stock in absolute ethanol made so that this addition was between 1 and 10 µL. Time courses of spectra were taken of this reaction to follow the kinetics and final yield. These were assessed using the A_{523} rather than the A_{500}, in agreement with published values for the λ_{max} of fluororhodopsin. The cuvettes were held in a water jacketed cell holder with temperatures set as indicated. When Ce6 was added, a mixture of Ce6 from a stock in DMSO as described above, and retinal in ethanol was made and added in place of the retinal.

2.2.3.2: PDS labelling of rhodopsin

Accessible cysteines in P23H rhodopsin were labelled by reaction with PDS. Rhodopsin in NaPi, pH 6 + 0.05% DM was diluted to 1.5 µM. PDS was added to the same buffer to 37.5 µM from a 0.1 M stock in ethanol. 300 µL PDS in buffer was made up to 400 µL with buffer in the reference cuvette, and with 1.5 µM rhodopsin (final rhodopsin 0.5 µM) in the sample cuvette. Time courses of spectra were recorded to follow the reaction at 20°C

2.2.3.3: MTSL labelling of rhodopsin

Rhodopsin was derivatized with S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) to label specific cysteines with a nitroxide spin label, which increases the relaxation of magnetization in NMR experiments. Through this it was hoped that distances between methionines and cysteines could be used to assign peaks in an NMR spectrum (see section 1.1.4.1 for discussion of relaxation and section 5.3.3 for the rationale behind MTSL labelling).

Accessible cysteines in rhodopsin were labelled with MTSL for the paramagnetic relaxation enhancement effects upon its NMR spectrum. The protocol for MTSL labelling (Resek et al., 1993) was adapted in two ways for different desired results. The first adaptation was to scale

up the labelling from ~200 µg to ~4 mg. This adaptation was based on the scaling up of 4-PDS labelling (Hwa et al., 1999) from 500 µg to 10 mg rhodopsin, as both protocols are used to label free cysteines. From the small to the large quantities of rhodopsin, there is a five-fold increase in rhodopsin concentration (and a change in volume) and a ten-fold increase in 4-PDS concentration. As the existing protocol for labelling small amounts of rhodopsin with MTSL used a 1:42.02 ratio of rhodopsin:MTSL, the new protocol used ~84.04 MTSL per rhodopsin.

Total beads containing 2 mg rhodopsin were then incubated with spin label (200 µM in 10 mL 5 mM MES). After 3 hours reaction at RT, the resin was washed with 30 bed volumes (24 mL) of 5 mM MES.

The second adaptation was also for a large-scale labelling, but attempting to label one of the two accessible cysteines, and without labelling on the column. The excess was reduced to 10X MTSL for this protocol, but was otherwise unchanged.

2.2.3.4: SDS denaturation

For both pSRII and bovine rhodopsin, SDS was added to denature the protein. A 30% (w/v) stock of SDS was made up in their respective buffers, and kept in a 25°C water bath to keep the SDS from precipitating. SDS was added to the cuvette in the dark to the desired concentration and mixed end over end. The time of addition was noted, and a time course of spectra was recorded to follow the unfolding reaction. Notes on the decomposition of spectra can be found in sections 2.3.2.1 and 4.3.2.

SDS denaturation with hydroxylamine

pSRII was denatured as with SDS above, but hydroxylamine hydrochloride (Sigma Aldrich) was added to 10 mM.

2.2.3.5: PNGase F treatment

Recombinant rhodopsin displays heterogeneous glycosylation, the pattern of which differs in some ADRP mutants (Kaushal et al., 1994). PNGase F is an effective and non-specific enzyme which removes N-linked glycosylation. By treating rhodopsin and ADRP mutants with PNGase F, the sources of different weights of bands in western blots was examined.

The concentrations of WT and N15S and P23H mutant rhodopsin samples were determined by absorbance spectroscopy, and diluted to match concentrations. PNGase F treatment was carried out according to manufacturer's instructions.

2.2.3.6: Western blots

Acrylamide gels were prepared with the stacking gel at 4% and the resolving gel at 10%. PNGase F treated rhodopsin samples were mixed with Laemmli buffer (Laemmli, 1970) and added to sample wells, along with a marker lane. A voltage of 5-10 V/cm gel was applied for around 10 minutes until the dye front reached the resolving gel, then a voltage of 120 V applied until the dye front neared the end of the gel.

The gel was equilibrated in ice cold transfer buffer with two sponges and Whatman paper. A polyvinylidene fluoride membrane (Thermo scientific) was wetted in methanol for 30 seconds, then incubated in transfer buffer. These were assembled into a sandwich and clipped into a gel holder cassette. This was placed into an electrode assembly and a buffer tank, which was then filled with transfer buffer. A current of 250 mA was applied for 70 minutes at 4^{o}C.

The membrane was incubated in blocking solution for 1 hour at room temperature, and in a solution of the 1D4 antibody, raised against the C-terminal nonapeptide of rhodopsin, overnight at 4^{o}C. The antibody solution was recovered for up to 3 times of reuse and the membrane washed in TBS-T three times for 20 minutes each, then twice in distilled water for 20 minutes. A solution of anti-mouse antibody (to detect the bound 1D4 antibody) was added. The antibody is joined to horseradish peroxidase, which stains purple when 4-chloro-1-naphthol and hydrogen peroxide is added in the developing solution. The developing solution was added and development allowed for ~3 minutes, then washed with tap water.

2.2.4: NMR of rhodopsin

2.2.4.1: 1- and 2D NMR of ^{13}C, ^{15}N-methionine rhodopsin

All protein samples containing [$^{13}C_5$, ^{15}N]methionine rhodopsin were placed in Shigemi NMR tubes. 1D ^{1}H NMR spectra were measured using a Bruker Avance DRX800 equipped with a 5 mm TXI CryoProbe (HCN/z). 1D ^{1}H spectra were recorded with 512 complex data points (acquisition time = 1 s). Water suppression is achieved using a 3-9-19 WATERGATE sequence in combination with gradient pulses.

2D {^{1}H, ^{13}C} HMQC spectra were measured using either the Bruker AvanceIII AV600 spectrometer with 512 × 76 complex data points ($t_{1,max}$ = 25.2 ms) or the Bruker Avance DRX800 with 512 × 100 data points ($t_{1,max}$ = 25.0 ms). All spectra were recorded with 256 scans unless indicated otherwise.

2.2.4.2: NMR of rhodopsin reconstituted with 14-fluororetinal

A sample of rhodopsin in NaPi buffer with ~1% DM (depending on the original volume concentrated into an NMR sample) was diluted into two 300 µL volumes of ~47.6 µM in Eppendorf tubes. The samples were reconstituted with a five-fold excess of 14-fluororetinal as described above, keeping the samples in a 4°C waterbath for 50 minutes, as the yield reached its maximum by this time in previous attempts. One sample was reconstituted in the presence of 10-fold excess Ce6. The samples were transferred to Shigemi tubes (Sigma Aldrich).

1D ^{19}F NMR spectra were measured using a Bruker AvanceIII AV600 spectrometer equipped with a 5 mm QCI CryoProbe (HCNF/z). All ^{19}F NMR chemical shifts were referenced to $CFCl_3$ (trichloro-fluoro-methane) as external standard. All spectra were recorded with 4997 complex data points (acquisition time = 1 s) and without broadband proton decoupling. Comparison against spectra recorded with broadband proton decoupling showed no differences (data not shown). Water suppression was achieved using a 3-9-19 WATERGATE sequence in combination with gradient pulses.

2.2.5: pSRII biophysical experiments

2.2.5.1: Collection of fluorescence timecourses of pSRII unfolding

Fluorescence spectra were measured at 25 °C using a LS55 fluorescence spectrometer (Perkin Elmer) (excitation slit: 5.0 nm, emission slit: 17.5 nm). Excitation was at 295 nm, and emission was measured at 335 nm and an integration time of 3 s. The typical experimental dead-time was 25 s.

2.2.5.2: Collection of circular dichroism timecourses of pSRII unfolding

Circular dichroism (CD) in the far UV reflects changes in secondary structure of a protein. Ellipticity of pSRII samples was recorded at fixed wavelengths of 222 nm, the wavelength at which maximum signal for helix content is found, at 30-second intervals for 2 hours, using an Aviv 410 spectrometer (Aviv Biomedical Inc.) and a 1×1 mm cuvette. The typical experimental dead-time was 1 min.

2.3: Data Analysis

All data analysis was performed in R (R Core Team, 2015), and non-linear least squares fitting performed with the minpack.lm package (Elzhov et al., 2016) unless otherwise stated.

2.3.1: Conservation of rhodopsin

2.3.1.1: Selection of residues for analysis

2.3.1.1.1: G-protein binding

To probe the interference of Ce6 with the G-protein binding, the conservation of relevant residues was investigated. Candidate residues were selected using the G-protein bound crystal structure of β_2 adrenergic receptor (β_2AR). All residues within 5 Å of any G-protein subunit were candidates for analysis. The β_2AR primary sequence was aligned against *B. Taurus* rhodopsin using clustal ω (Noorwez et al., 2004; Saliba et al., 2002). 29 of the 33 candidates aligned with rhodopsin.

2.3.1.1.2: Ce6 binding

A putative Ce6 binding site was identified with docking in (Woods et al., 2017). These residues were used as the Ce6 binding site.

2.3.1.1.3: Retinal binding

Using the 1L9H rhodopsin crystal structure, all residues within 5Å of retinal were selected.

2.3.1.2: ConSurf

The sequence of bovine rhodopsin was submitted to the ConSurf server, and set to search the "clean Uniprot" database with default settings to collect 150 sequences that sample the list of homologues. The resulting multiple sequence alignment (MSA) was downloaded and the partial rhodopsin sequence from *M. niger* was added and aligned to the MSA using clustal ω. This was resubmitted to the ConSurf server. Part of ConSurf's output capability is an annotation of the original uploaded sequence indicating the relative conservation at each position which can be used to colour the PyMol model of the protein. This was used to colour the G-protein interface of bovine rhodopsin (1L9H) with the colour-blind friendly colour key.

2.3.1.3: Conservation plots

The 151 sequence MSA from ConSurf was imported into R, as well as the ConSurf output. R scripts were written so that plots could be made indicating the percentage identity to the most common residue for specific positions, and how specific sequences compared with the consensus at these positions. The consensus sequence used is the most common residue letter code found at that position. The average conservation of the consensus is calculated as the mean of the count of this code divided by the total number of sequences with a read at this position.

When referring to specific positions, the user specifies a reference sequence within the set and positions in that sequence. In these results, the *B. taurus* sequence was used as a reference, but in future any sequence within the alignment could be used. The script removes any gaps in the reference sequence caused by alignment and retrieves the positions specified as they are in the potentially gappy alignment. The user also specifies a set of sequences to compare with the reference and the consensus. These are entered as strings which must match a sequence name in the alignment, which the script uses to retrieve the letter codes (or gaps) at the specified positions, counts the instances of that code at that position and divides that count by the total sequences to deliver a fraction. The fraction of sequences matching consensus are plotted as a grey bar of width matching the number of query sequences. The fraction of sequences matching the query sequences are then plotted as coloured bars on top. The mean fraction matching the consensus is plotted as a dotted line across the plot.

The residues at the query positions are also displayed as a table underneath the bars. The possible combinations of matches/mismatches are calculated and a colour palette with that many colours is generated and assigned to the possibilities. If all sequences match the consensus, no colour is produced. Cells with mismatches in any sequence are coloured accordingly, e.g. if the first sequence matches but a second does not, one colour is used; and if the first does not match but the second does, another colour is used etc.

2.3.2: Unfolding kinetics

2.3.2.1: Absorbance

2.3.2.1.1: Consecutive reactions

Individual spectra of unfolding time points were imported as comma separated values (csv) files and compiled into a matrix, with each spectrum making up a column. The first column of this matrix is the wavelengths of recorded values. A second matrix of guesses for parameters (extinction coefficient, width and centre for peaks as described in 4.3.2) for each of the peaks is provided by the user. To further define the model, a Boolean value for whether a scattering baseline should be added is provided, as well as a numerical vector defining which rows of the matrix of guesses refer to peaks which remain static throughout (in all examples, a single peak for the 280nm absorbance). Further information about the sample is supplied by the user as a numerical vector of the times for each spectrum in minutes, numerical values for the concentrations of pSRII and SDS, a numerical value for the wavelength cutoff (used early on when terms for the scattering and static peaks were not

included and maintained in later versions for completeness), and a numerical value for the repeat, as experiments were performed in triplicate.

If any static peaks were included in the model, the first spectrum is fit by NLS as the sum of all peaks in the initial guesses. A function within the script generates models for any number of peaks, with or without a baseline as defined by the user. The fitted values for the static peaks are recorded. If a wavelength cutoff is supplied by the user, rows of the matrix of spectra with wavelengths below this value are discarded.

The fitting model used for the time course is then defined. The mixed kinetic/Gaussian terms are augmented with the scattering and static terms as required, yielding a formula like equation 7 in chapter 4. The rationale for mixing kinetic and peak fitting terms is also explored in chapter 4.

The matrix of spectra is then unfolded into a three-column matrix in which the first column is the vector of wavelengths of recorded values, taken from the original matrix, repeated as many times as there are timepoints; the second is the time of recording each value, requiring each value is repeated as many times as there are recorded wavelengths. The last column is the corresponding absorbance value recorded at each timepoint and wavelength. This matrix is fitted to the model, using the supplied guesses as starting parameters. Limits are applied so that no peak can be negative and that no peak's centre will fall outside the recorded spectrum.

Once fitting has taken place, the data and fitted peaks can be plotted. The plot area is divided in two, the left for the absorbance spectrum, and the right for the progress of the reaction. Each plot is for a single time point. The spectrum for that time point is plotted as points first, with the y-axis limits fixed to the minimum and maximum absorbance for the whole data set so that all plots are scaled the same. A line for the predictions of the total model is then plotted as a dashed line, followed by coloured curves for the scattering baseline and each of the fitted peaks, then the residual for that timepoint is plotted as a finely dotted line. The right side has axes plotted between 0 and 1 in the y-axis and 0 to the last timepoint recorded in the x axis. Fitted curves for the relative concentrations of the three components described in the kinetic model (chapter 4) are plotted with colours corresponding to their curves in the fitting on the left side. A dotted line is added, indicating the time.

The script returns the fitting as an NLS object, the fitted coefficients in a separate list for ease of access, and can optionally output plots either as a .pdf (useful for producing vector

graphics which can be separated and rescaled freely) or a .jpg (useful for producing animations).

2.3.2.1.2: Single reaction

Refolding pSRII and pSRII unfolded in the presence of hydroxylamine were modelled as single reactions between native pSRII and free retinal as described in chapter 4. Spectra were imported as for the consecutive reactions. Each spectrum was fitted to a model combining a scattering contribution, a static protein peak and the retinal contributions. These contributions were divided into the native profile and free retinal as described in chapter 4. A function fitting the data to this model is applied to each column of the matrix. The amplitudes of the retinal peaks in these fits are put in one column of a data frame, and the multiplier of the native profile ("B" in the model) in the other. The data frame is fit to equation 13 in chapter 4, as the two should account for all retinal present. As one increases, the other will decrease, but by how much is unknown. By fitting the two to a constant, this ratio can be calculated. A data frame containing the concentrations at each timepoint, as calculated from this, is constructed, with another column containing the time points. The concentration of free retinal is fitted to an exponential growth decay model. The function returns a list containing the data frame of concentrations, the fitted models for all spectra, the concentration fitting, and a figure for the fraction of pSRII permanently unfolded. This is estimated by taking the asymptote ("C" in the model, chapter 4, equation 15) and dividing it by the total concentration of the sample, supplied by the user.

Optionally, similar plots to those for consecutive reactions can be returned.

2.3.2.2: Fluorescence and Circular dichroism

The changes in fluorescence emission at 340nm and in molar residue ellipticity (MRE) at 222nm for unfolding pSRII were tracked over time. The data were imported into R as two column matrices, one column for time and the other for the recorded value. The minimum recorded value for a timecourse was subtracted from the second column (rationale can be found in chapter 4). Initial parameters were not user supplied but found using a self-starting algorithm. This assumes the data follows a single exponential growth curve with an amplitude A, meaning the maximum recorded value can be used as a starting parameter for A. The rate constant of an exponential decay is approximated by $\frac{\ln(2)}{t_{1/2}}$, where $t_{1/2}$ is the half-life. The recorded value closest to half of the maximum was used for the half-life.

The data is then fitted to an exponential growth decay curve as described in chapter 4. The self-starting algorithm is then applied to the residual. The fitted values and the guesses for the residual are used as starting parameters for a double exponential model as described in chapter 4. A third column is added to the matrix which shows only the user supplied concentration for pSRII in the sample. If the second fitting is successful, the fitted values are used (with the amplitudes divided by the concentration) to fit a combination of kinetic terms described in chapter 4, which can also be expressed as a superposition of two exponential growth decay curves:

$$Fluorescence \sim [A]_0\left(\frac{k_1}{k_2-k_1}\left(e^{-k_1 t}-e^{-k_2 t}\right)\right) + [A]_0\left(1+\frac{1}{k_1-k_2}\left(k_2 e^{-k_1 t}-k_1 e^{-k_2 t}\right)\right)$$

Bracket out $[A]_0$

$$Fluorescence \sim [A]_0\left(\left(\frac{k_1}{k_2-k_1}\left(e^{-k_1 t}-e^{-k_2 t}\right)\right) + \left(1+\frac{1}{k_1-k_2}\left(k_2 e^{-k_1 t}-k_1 e^{-k_2 t}\right)\right)\right)$$

Remove inner brackets

$$Fluorescence \sim [A]_0\left(\frac{k_1}{k_2-k_1}\left(e^{-k_1 t}-e^{-k_2 t}\right)+1+\frac{1}{k_1-k_2}\left(k_2 e^{-k_1 t}-k_1 e^{-k_2 t}\right)\right)$$

Expand brackets

$$Fluorescence \sim [A]_0\left(\frac{k_1}{k_2-k_1}e^{-k_1 t}-\frac{k_1}{k_2-k_1}e^{-k_2 t}+1+\frac{k_2}{k_1-k_2}e^{-k_1 t} - \frac{k_1}{k_1-k_2}e^{-k_2 t}\right)$$

Collect terms

$$Fluorescence \sim [A]_0\left(1+\frac{k_1}{k_2-k_1}e^{-k_1 t}+\frac{k_2}{k_1-k_2}e^{-k_1 t}-\frac{k_1}{k_1-k_2}e^{-k_2 t} - \frac{k_1}{k_2-k_1}e^{-k_2 t}\right)$$

Bracket out exponentials

$$Fluorescence \sim [A]_0 \left(1 + \left(\frac{k_1}{k_2 - k_1} + \frac{k_2}{k_1 - k_2}\right) e^{-k_1 t} - (\frac{k_1}{k_1 - k_2} - \frac{k_1}{k_2 - k_1}) e^{-k_2 t}\right)$$

Expand

$$Fluorescence \sim [A]_0 + [A]_0 \left(\frac{k_1}{k_2 - k_1} + \frac{k_2}{k_1 - k_2}\right) e^{-k_1 t} - [A]_0 (\frac{k_1}{k_1 - k_2} - \frac{k_1}{k_2 - k_1}) e^{-k_2 t}$$

The LM algorithm for fitting is tolerant enough of poor starting parameters that no other conversion has been necessary to use the parameters for the two-curve model for fitting the kinetic model.

At this point, the single exponential model is compared with the kinetic model to test which is more appropriate. When using Akaike's "An Information criterion" (AIC), functions built into base R are sufficient. The fitting with the lower AIC is used.

Packages in R do exist allowing cross validation, but the process is simple enough that 10-fold cross-validation was written directly into the script. Briefly, the data is trimmed to be divisible into ten evenly, removing random data. An additional column is added, placing the numbers one to ten randomly an equal number of times in this column. This divides the data into ten "folds". For each fold, the procedure is as follows: each fold is used as a test set. All other folds are the training set. The training set is used as the data for fitting to the model under examination. The predicted values for each value in the test set are calculated and the residual sum of squares between the predicted and actual values for the test set are recorded. Once this has been performed on all folds, the mean residual sum of squares for that model is calculated. This is performed on both the single exponential and kinetic models. The model with the lowest mean residual sum of squares is selected as the preferred model.

The data and the fitted values of the preferred model are then plotted with the residual.

2.3.3: Automatic peak picking

A common method of smoothing data is to calculate each point as a moving average of some window to either side of that point. However, if too broad a window is used, the spectrum becomes distorted, especially at peaks. While this is not desirable in displaying data, or making calculations, it does act as a discriminant of peaks.

A vector of the length of the original data minus the averaging window side was filled with the average of the window size from the same position in the original data. A disadvantage of using a large averaging window is that a region the size of half the window is lost from either end of the spectrum, so any peaks within this window will not be estimated in the

early stages. The smoothed values were subtracted from the original values from half the window into the data until half the window from the end to reflect this.

When using the oversmoothing strategy, an issue is that, because it works by averaging, half of the points are negative (chapter 4, figure 4b). This means that although small peaks register, they are not necessarily even local maxima. In image analysis there is a class of filters called geometrical filters, which simulate the reaction of some shape to features in the data. A useful background subtraction can be achieved by the rolling ball filter (Sternberg, 1983), which establishes a background accessible by an imaginary ball being rolled over a surface of the intensity values in an image. A rolling disc filter for two dimensions was written for the difference between the oversmoothed and original data, which rolls a disc along the bottom of the spectrum and brings the points the disc touches to zero, making all peaks local maxima between zeroes (Chapter 4, figure 4b (II)).

The filter calculates the point at which an imaginary circle touches the data at every point. As the data are discrete, this is actually half of an imaginary regular polygon with a vertex at each value of x within the radius chosen by the user. The difference in y between the centre of the polygon and each vertex is calculated. For each value of the data within the radius of the point in question, the centre of the polygon as it would touch this point is calculated. The lowest of these values is equivalent to where the imaginary disc would touch the data, and recorded as the baseline of the filter. The filter requires the radius of the circle and a scaling factor between the values of the x and y values. This baseline is subtracted from the data, pushing all values above or equal to zero.

After the rolling disc has been applied, a list is made of all regions above zero, recording the width of the region and the location of the local maximum. This is achieved by listing all the places where the filtered data is zero, then recording the mean and difference between values. The value of the original spectrum at that local maximum is added to the list (chapter 4, figure 4a). These values are used as starting parameters for the next stage.

This stage refines the peaks. For each peak, a region of the original spectrum two widths wide around the centre is fitted, using the parameters from the previous stage. This region is fitted to a Gaussian peak using the previous stage to provide guesses. The predicted values from this fitting are subtracted from the whole spectrum.

To ensure no peaks are missed, fitting is carried out until all residual values have an absolute value less than a predefined threshold, currently set arbitrarily by the user. The areas of the refined peaks are calculated and used as rankings for the next stage.

Fits are made to the whole dataset, adding peaks in order of their area. As each peak is added, the AIC between the model with and without that peak is made. This is repeated until model selection indicates that further peaks are extraneous.

2.3.4: Image analysis

NMR spectra were unfolded in the MatNMR toolbox for MatLab by collaborators at Cambridge, and exported as matrices which were opened in R. The 151x97 matrices were then exported as 97x151 images (Figure 4.15) with the ^{1}H data along the x axis and the ^{15}N data along the y axis to resemble the standard display of NMR spectra. Each pixel of the image represents the intensity of the spectrum at that point.

To identify features in this image, BLOB (Binary Large Object) detection was implemented. This is a relatively simple object recognition technique in image analysis. Pixels of interest in the image are separated by some criterion from the background and grouped into objects by a filter. In this implementation, the discriminant was threshold on intensity applied, where all values below the threshold are false, and all above it are true (chapter 4, figure 9b). The threshold is supplied as a fraction of 1 by the user. Continuous areas of true values are labelled as a blob. Blobs are assumed to be roughly circular and the centre and radius are recorded (chapter 4, figure 9c). This method is somewhat vulnerable to picking up noise, so all blobs with a radius smaller than a threshold can be filtered out. With the low resolution of the bucketed spectra used this was of limited use, but higher resolution spectra may benefit from such a filter.

2.3.5: Decomposition of Ce6 + rhodopsin spectra

Spectra of Ce6 and rhodopsin mixtures were imported from csv files and collated into matrices as for pSRII unfolding time courses. The first spectrum was a pure rhodopsin spectrum, and the second an unilluminated rhodopsin/Ce6 mixture, for which the volume of Ce6 added had been recorded. The change in rhodopsin concentration from volume change was corrected by multiplying the initial spectrum by $400/(400+V_{chlorin})$. The chlorin contribution was calculated by subtracting the initial spectrum from the second and was appended to the matrix.

A vector of the ratios of the 664nm value of each of the original Ce6/rhodopsin spectra was divided by the 664nm value in the extracted chlorin spectrum. To each of these spectra, a function was applied which subtracted the chlorin spectrum multiplied by the relevant value from the ratio vector. This accounted for the chlorin contribution after any bleaching, assuming uniform bleaching over all wavelengths. The shortcomings of this approach are discussed in chapter 3.

The A_{500} values for each of these chlorin-subtracted spectra were stored in a vector, along with the 660nm value for the first spectrum containing Ce6. The change in A_{500} was fitted to an exponential decay model, and the rate constant was plotted against the concentration as estimated by using the 660nm value and an extinction coefficient of 59000 (Oseroff et al., 1986).

2.3.6: Decomposition of Western blot data

The pixel density across lanes was extracted using imageJ (Schneider et al., 2012). To ensure the lanes were aligned in later stages, a pure white line was drawn across the top of each blot. The csv files produced were imported into R, and any values found outside the two values reading 255 (the maximum for 8-bit greyscale) were discarded. The values were then inverted so 0 became 255, 1 became 245 etc. and scaled between 0 and 1. The peaks caused by the ladder were automatically picked using a Savitzky-Golay filter to estimate the second derivative (Savitzky and Golay, 1964). This operation was performed, then the local minima within a user defined window were found, and ranked by magnitude. The user supplied a vector of the ladder molecular weights (MW), and the top n (where n is the length of this vector) local minima were assigned as the ladder positions.

A linear regression was performed on the log-values of the pixel position of each peak against the log-values of the theoretical MW. This regression was used to estimate a MW for each pixel value of blot profiles.

Four blots were prepared in this way and collated for analysis. As the acquired images had regions in which the signal was saturated, each band was assumed, in the pixel density profile, to have a normal distribution. The lanes were separated into a list of matrices in which the first column is MW and the second intensity at that position. Any row where intensity was greater than 0.95 was removed so all values above 95% density for each lane were discarded as saturated. For most blots, the profile above 50 kDa was largely discarded in this way, so further analysis was limited to the region between 20 and 50 kDa. The background signal was then corrected for by using the rolling disc filter as described for

automatic peak picking above. Gaussian peaks were then fitted by non-linear least squares to the remaining data, and the peak centres (in kDa) and share of the total area occupied by each was recorded. Boxplots summarising this information were constructed, with band centre in the y-axis and relative width of boxes representing fraction of total area occupied.

2.3.7: PDS time courses with P23H

The reaction of 4-PDS with cysteine results in a peak centred at 323nm. To follow this reaction and the bleaching observed with P23H, the spectra were decomposed into a scattering baseline, a protein peak at 280nm, a 323nm peak for 4-PD, a 380nm peak for retinal, and 500nm peak for P23H. Unlike for SDS unfolding, this reaction had no explicit kinetic model, so this could not be used to use the entire dataset to decide on peak parameters. Instead, each spectrum was fitted separately, and the parameters (amplitude, centre and width) of the peaks over time were plotted. As the area of the rhodopsin peak decreases significantly within the first few minutes, the fitted centre and width for the initial spectrum was taken and fixed as the centre and width of that peak and other parameters were allowed to vary. The parameters of the other peaks were plotted, and the timepoints for which the centres and widths were stable were judged by eye. The means of these parameters were taken and fixed. The amplitudes were allowed to vary, and these were taken as the absorbances of the different species present.

The centre of the P23H rhodopsin peak was found to be at 476nm, similar to previously published values for P23H (Noorwez et al., 2004). The extinction coefficient of this was assumed to be the same as WT rhodopsin's at 500nm. The ε_{323} of 4-PD is 19000 $cm^{-1}M^{-1}$, and was used to calculate the concentration of 4-PD at a given time. This was divided by the concentration of rhodopsin present to find the number of cysteines reacted. This was fitted to a single exponential decay curve and a superposition of two exponential decay curves, and the AIC was used to compare the models of labelling kinetics.

CHAPTER 3: RHODOPSIN AND VISION

3.1: Summary

Chlorin e6 (Ce6) has been shown to enhance bleaching of rhodopsin by far-red illumination by an unknown mechanism, which indicates a possible role for chlorophyll derivatives found in the eyes of dragonfish . Previous characterization of this enhancement has made no allowance for the bleaching of Ce6 itself, and has left the mechanism unclear. To address this gap, systematic bleaching experiments have been carried out, which, coupled with effective decomposition methods, have allowed accurate analysis of the kinetics and spectral features of rhodopsin bleaching enhancement by Ce6. The effects are shown to increase with greater Ce6 concentration, and to be strictly light dependent. The subtraction method used enables the examination of higher ratios of Ce6:rhodopsin. The G-protein binding activity of bovine rhodopsin is inhibited by the addition of Ce6, so the conservation of the G-protein binding and putative Ce6 binding sites were investigated using the ConSurf server, and new visualisations of sequence comparisons for specific residues for selected species within a larger dataset were devised. The binding sites were found to be highly conserved both across the dataset and between bovine and *M. niger* rhodopsins. To probe specific interactions between retinal and Ce6, rhodopsin was reconstituted with 14-fluororetinal, and an in-cuvette reconstitution protocol optimised for ^{19}F NMR. The addition of Ce6 to fluororhodopsin made no significant differences to the retinal environment as detected by NMR.

3.2: Introduction

Porphyrins absorb strongly both in the UV and visible to near infrared (Figure 3.1a). Their spectra arise from the 18-electron aromatic system in the macrocycle of four joined pyrrole units (Gouterman, 1961). Peaks arising from transitions in this system can be divided into the strongly allowed transitions of the B-band around 400 nm and the weakly allowed transitions in the Q-band from 480-700 nm. The transitions leading to the Q-band are “forbidden” transitions in porphyrins, but are weakly allowed by molecular vibrations in the macrocycle. Chlorins are porphyrins in which one of the pyrrole rings has been reduced. This forces asymmetries upon the macrocycle, making the orbitals involved separate in energy, unlike their degeneracy in porphyrins. This increases the intensity of the peaks of the Q-band, and makes the furthest red transition more intense than the others. This makes chlorins appear dark green, unlike the red or magenta common in porphyrins. Further modifications are made to spectra by substitutions at different positions around the macrocycle, and

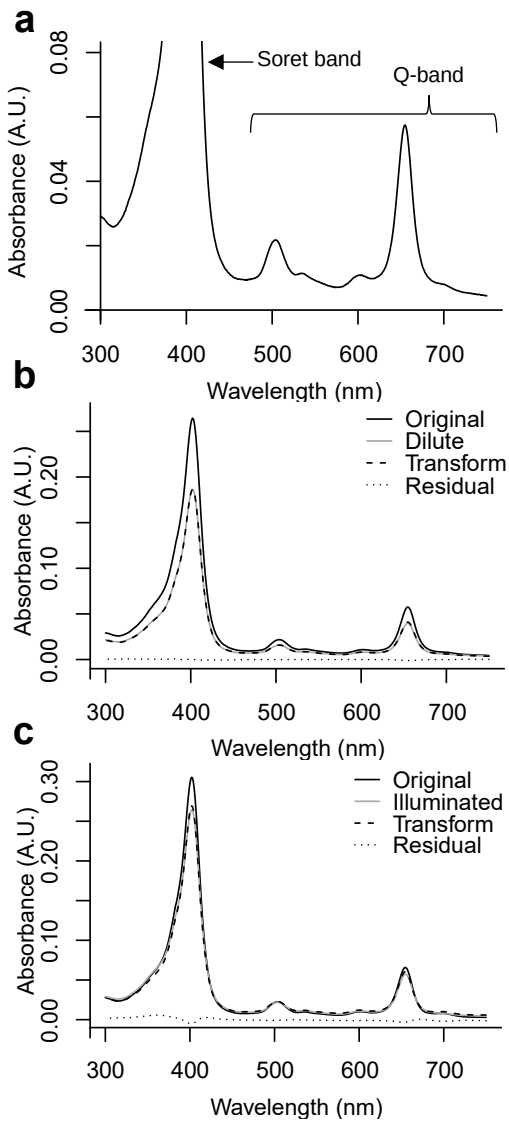


Figure 3.1: Spectra of chlorin e6 (Ce6) a) An example spectrum of Ce6, showing the Soret or B-band and the peaks of the Q-band. As the B-band is much more intense than the peaks of the Q-band, the output has been truncated. **b)** An absorbance spectrum of chlorin was taken, then the sample diluted with NaPi buffer. The spectrum of the diluted sample was fitted as a multiplication of the original spectrum and the original and diluted spectra match. **c)** The same fitting procedure was followed after bleaching instead of dilution and mismatches can be seen in the residual, especially around 400 and 665nm

modifications making the core of the macrocycle more symmetrical, such as coordination of a metal ion or protonation, simplify the Q-band.

Previous biochemical investigation of enhancement of red light bleaching of rhodopsin has been limited (Ilyas Washington et al., 2004; Washington et al., 2007). Peaks of the Q-band of Ce6 overlap with rhodopsin's retinal peak, and the B-band overlaps the absorbance of free retinal. Ce6 itself bleaches a small but significant amount upon illumination, and Ce6's absorbance is stronger than rhodopsin's at the same concentration, so when Ce6 is added to a rhodopsin sample at increasing concentrations its bleaching can dominate changes in spectra. The goal of this work was to adapt methods described for decomposition of absorbance spectra used to determine unfolding kinetics in pSRII (Chapter 4) and determine whether they could better define the effects of Ce6 on rhodopsin by separating their contributions to absorbance. Ce6 seems to also inhibit the interaction between rhodopsin and its G-protein (Eric Gardner, Kalyan Tirupula and Judith Klein-Seetharaman, unpublished results).

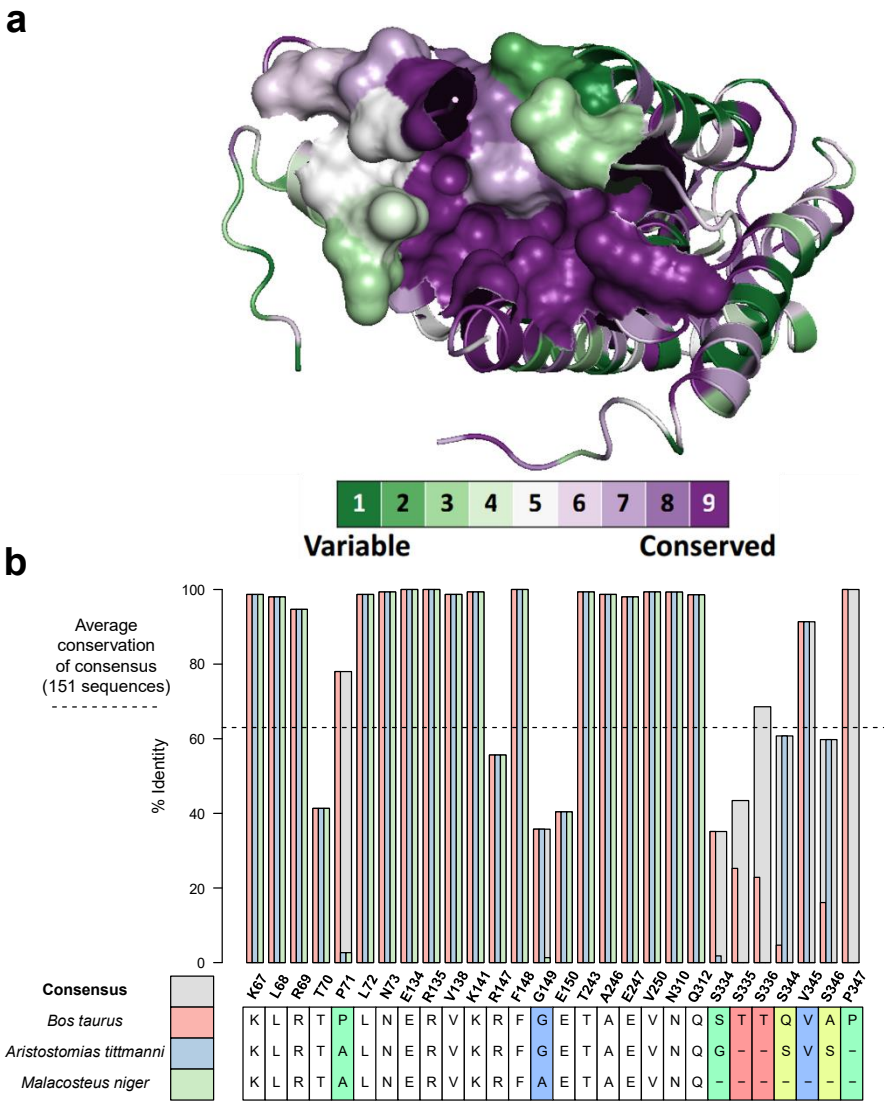


Figure 3.2: Consurf analysis: Interspecies conservation of putative Ce6 binding residues: (a) The 1L9H structure of rhodopsin coloured with the ConSurf scores. The Ce6 binding pocket identified by docking is represented as a surface, the rest in cartoon representation. **(b)** Conservation plots comparing 3 rhodopsin sequences with the consensus for 150 species with *M. niger* added. Residues identified for Ce6 binding, as described in methods.

3.3: Results

3.3.1: Conservation of Chlorin binding domain

3.3.1.1: ConSurf analysis of putative binding site

Chlorin binds in the cytoplasmic domain of rhodopsin (Woods et al., 2017). To better understand the potential functional significance of this finding, we conducted conservation analysis. Figure 3.2a shows the 1L9H structure of bovine rhodopsin coloured according to the relative conservation of residues. The Ce6 binding site from (Woods et al., 2017) is shown as a surface, and shows that, though some residues around the periphery of the binding site are poorly conserved, this region is highly conserved across a wide range of species.

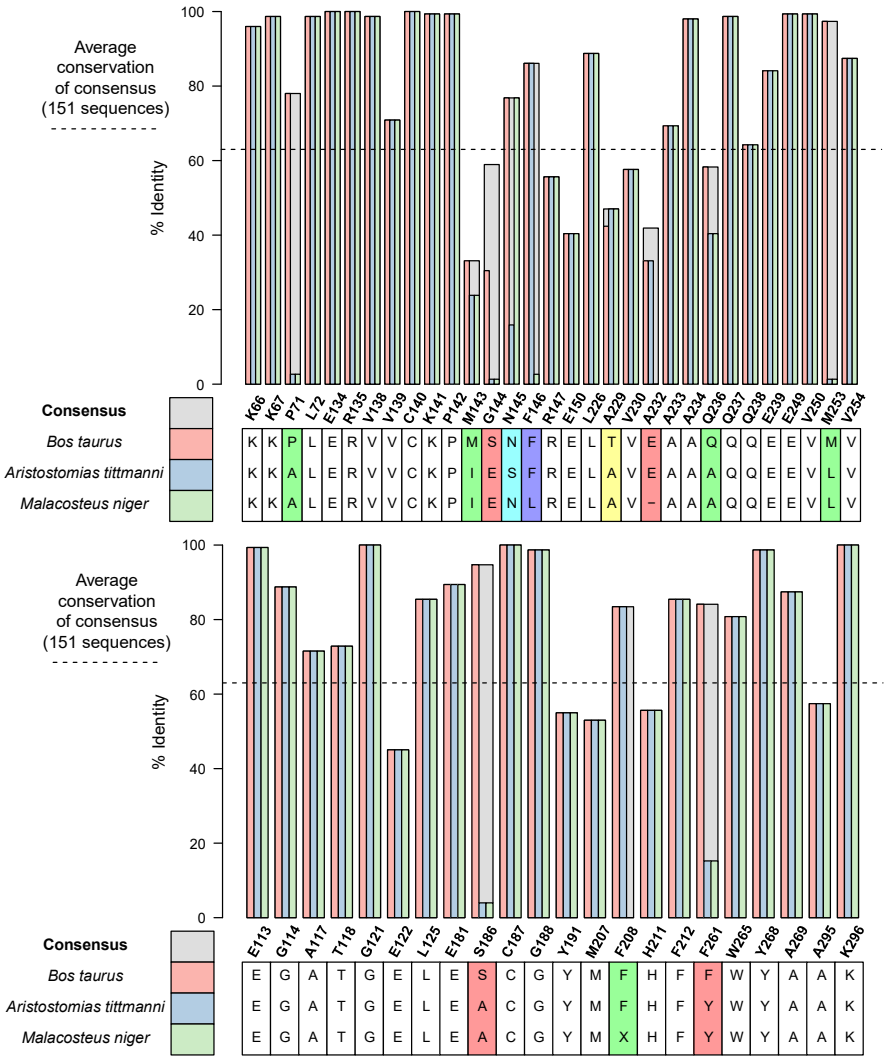


Figure 3.3: Conservation of other binding sites in rhodopsin. Plotted as in Figure 3.2b for **(a)** G protein and **(b)** retinal binding sites, as identified in methods

3.3.1.2: Conservation plots

Though the ConSurf coloured structure is useful to represent the relative conservation of residues, it does not show where *M. niger* is identical to mammalian rhodopsin. For this purpose, plots were devised to compare the same sites in different organisms to one another and the consensus sequence.

The MSA from ConSurf analysis was exported as a text file and imported into R. Scripts were written to compare any given sequence to the rest of the sequences at a given position. Building on this, a script was designed to graphically represent these comparisons. Examples are shown in Figure 3.2b and Figure 3.3. The percentage of sequences matching each sequence compared at the positions queried is shown as a coloured bar superimposed on a grey bar showing the percentage of sequences matching the most common residue at that position. Below the bars, a table shows the residues at the queried positions for each

sequence compared. Coloured columns in this table represent which of the compared sequences agree with consensus. Positions corresponding to predicted Ce6 and G-protein binding sites were compared between *B. taurus, M. niger,* and *Aristostomias tittmanni (A. tittmanni)*, a species within the same family (Stomiidae) as *M. niger* which was part of the original ConSurf MSA.

Rhodopsin is well conserved, especially in key residues identified as being of interest for Ce6. *B. taurus* and *M. niger* rhodopsins are identical for most residues in the putative Ce6 binding site (Figure 3.2b) and are also identical with the consensus sequence of 151 species. In seven positions, the sequences of both *A. tittmanni* and *M. niger* show gaps. This is not due to the MSA carried out, but only partial sequences being available for alignment. One potentially significant difference is the substitution of alanine for proline at position 71 in both *A. tittmanni* and *M. niger*. The proline is fairly well conserved at 78% across the dataset, and the secondary amine group has unique structural characteristics. P71 is found at the interface between cytoplasmic loop 1 and helix II. The N-terminus of an α-helix is a common position for prolines (Kim and Kang, 1999), and it is unlikely that the structure breaking characteristics of proline are in effect here.

31 residues were used to represent transducin (G_t) binding as described in section 2.3.1.1. Of these, seven had below average conservation (Figure 3.3a). Nine residues displayed a mismatch between the query sequences and consensus. All but two of these had below average conservation, one of these being P71. The other was M253. Though this residue is almost universally conserved in this dataset, it has been chosen in the past for labelled with azido groups specifically because it has no reported functional interactions (Ye et al., 2010). There is significant overlap between Ce6 and G_t binding sites, so it is possible to explain the conservation of the Ce6 binding site without a selection pressure maintaining Ce6 binding. However, the aim here was simply to show that the binding site does not differ and that conclusions about Ce6 binding based on bovine rhodopsin were transferable.

The retinal binding residues are less well conserved (Figure 3.3b) because the dataset includes differently tuned opsins. Five of the 22 residues are of below average conservation, but the query sequences agree with the consensus here. Three positions have a degree of mismatch, though F208 only mismatches due to a misread. The other two positions, S186 and F261 are replaced by A and Y in the fish respectively. Modifications to the 261 position result in a red spectral shift (Yokoyama et al., 1995), and specifically the F261A mutation is known to reduce G_t activation (Han et al., 1996) and affect early photolysis events (Jäger et

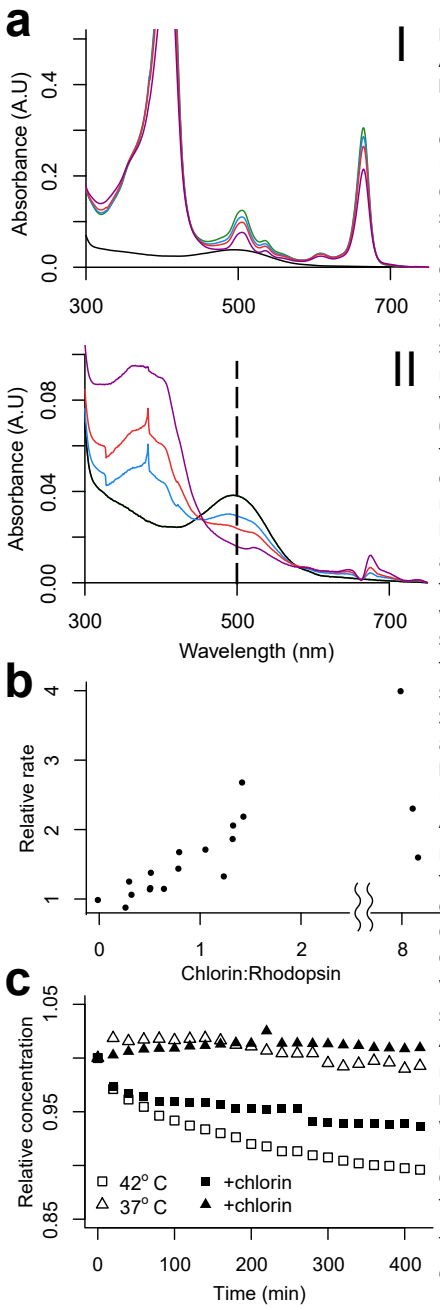


Figure 3.4: Absorbance spectroscopy. (a) Absorbance spectra of Rhodopsin bleaching by 630nm light in the presence of Ce6. (I): Raw spectra with chlorin added to half the concentration of rhodopsin. The initial Rhodopsin spectrum was multiplied by the dilution factor from adding Ce6 and subtracted from the spectrum with chlorin added to obtain a spectrum for the Ce6 contribution. Between each subsequent spectrum, the sample was illuminated through a 630nm long-pass filter for 30 seconds. A spectrum was taken after exposure to the room's ambient light to ensure all bleaching was caused by the light source. **(II)**: Subtraction of the Ce6 contribution from the bleaching time course. The Ce6 contribution was calculated by multiplying the reference spectrum acquired by subtraction by the ratio between the A_{665} of the reference and that timepoint. The dashed line running through 500nm is shown to indicate that without the Ce6 contribution the rhodopsin spectrum is not altered. The black line shows the initial spectrum. The green line in a(I) shows the first spectrum with chlorin added. Spectra at 60, 120 and 300 seconds are shown as blue, red, and purple lines respectively. **(b) Enhancement of bleaching rates by chlorin.** Remaining rhodopsin was estimated using the A_{500} of chlorin subtracted spectra. These were normalised between the initial spectrum and the 495nm bleached spectrum to calculate the extent of bleaching, and single exponential decay curves were fitted to the timepoints to estimate rate constants. Chlorin concentration was estimated using the A_{664} of the initial Ce6 spectrum, and rhodopsin from the range of A_{500} values. Rates are shown relative to the rate of bleaching with DMSO alone. **(c)Dark noise in the presence of chlorin.** Rhodopsin was kept at the temperatures indicated in the presence and absence of ~5X chlorin. Chlorin's contribution was subtracted as described, and the dark noise estimated using the A_{500} at each timepoint. Similar results to 37°C were observed at 25°C but omitted for clarity.

al., 1997). S186 is known as a spectral tuning residue (Sekharan et al., 2013; Yokoyama et al., 2007) and important for the internal hydrogen bonding network of rhodopsin (Okada et al., 2002), though it should be noted the other residues important for this network are conserved between *B. taurus* and the fishes.

Overall, the conservation of the residues relevant for the interaction between Ce6 and rhodopsin is high. Differences between *B. taurus* and *M. niger* specifically are minor and seem unlikely to have any functional relevance. The differences in the retinal binding pockets between these species are consistent with a red-shifted absorbance.

3.3.2: Bleaching experiments

To investigate the effects of Ce6 upon the red sensitivity of rhodopsin, mixtures of rhodopsin and Ce6 were illuminated with light with wavelengths below 630 nm filtered out for 30 seconds at a time and measured after each illumination. After five minutes of illumination total, the sample was illuminated through a 495 nm filter. Normally rhodopsin is not reactive to the red light through the 630 nm filter but will be quickly bleached by the light through the 495 nm filter.

3.3.2.1: Ce6 subtraction method

To properly assess the effects of Ce6 upon rhodopsin bleaching, a protocol to subtract Ce6's spectrum from the overall spectrum was devised. It is difficult to reliably add equal Ce6 to both the reference and sample cuvette, and previous investigations simply ignored any changes to Ce6. With high relative concentrations of Ce6, any changes in the rhodopsin sequence will be bleached by small changes in the Ce6 spectrum, so the decomposition method made mixtures with high Ce6 accessible to analysis, while making it possible to analyse the kinetics of rhodopsin bleaching separately from any Ce6 changes.

To decompose spectra, an initial, rhodopsin only, spectrum was measured. Then a known volume of Ce6 stock was added, and a spectrum measured. The initial spectrum was corrected for the change in rhodopsin concentration and this spectrum was subtracted from the second to isolate the Ce6 contribution. The 664 nm peak was unaffected by rhodopsin, so this spectrum was multiplied by the ratio of the A_{664} for any mixed spectrum to the calculated Ce6 spectrum to subtract its contribution and account for its bleaching. Generally, this successfully yielded normal looking rhodopsin spectra (Figure 3.4a II), with small amounts of deviation in the far red, and increased noise around 400nm, presumably due to the high absorbance of chlorin in that range.

3.3.2.2: Ambient light

To check that the bleaching of rhodopsin in the presence of Ce6 was solely due to illumination with the filtered light, samples at 1 and 0.5:1 Ce6 were treated as illuminated samples, but with the fibre optic covered (Figure 3.5a). This was to check both that ambient light, such as unfiltered light from the light source, was not interfering and that no light-independent effects of Ce6 were relevant on the experimental timescale. At the lower Ce6 concentration there is no discernible difference between before and after 30 seconds of ambient light. At 1:1 Ce6, there is some small amount of change in the spectrum resembling increased scattering, but rhodopsin seems stable.

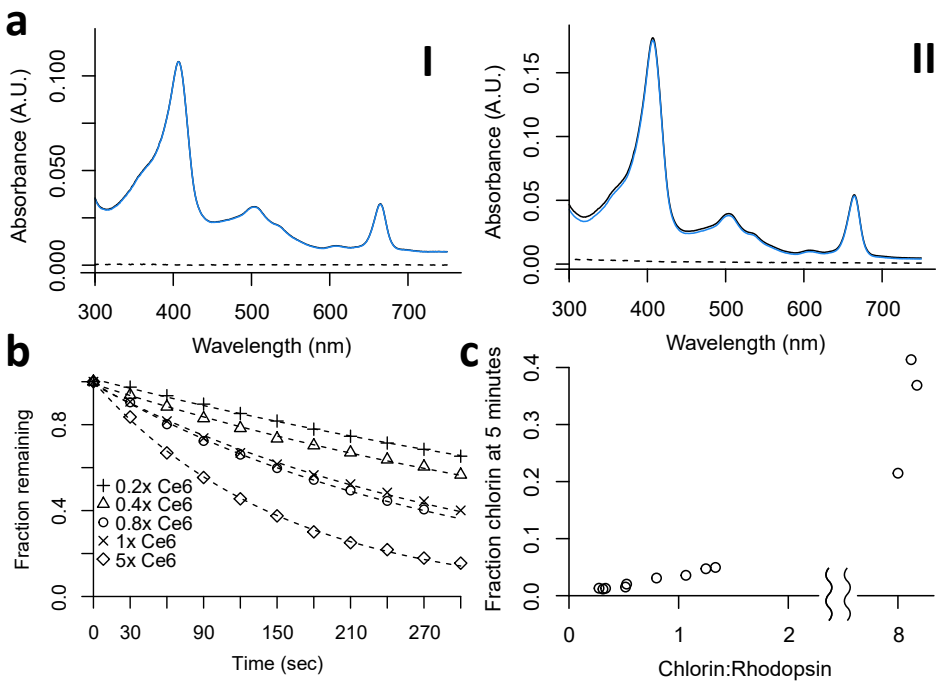


Figure 3.5: Other aspects of analysis of the bleaching of Rhodopsin in the presence of Ce6. a) Bleaching without illumination to test the effects of ambient light on rhodopsin with **I)** 0.5:1 and **II)** 1:1 Ce6:rhodopsin. Spectra before (black) and after (blue) exposure to ambient light are shown, though the spectra are similar, and the first spectrum is not visible in I. The difference between the two is shown as a dashed line. **b)** Example timecourses of bleaching of rhodopsin in the presence of Ce6 with illustrative fitting curves. **c)** Bleaching of the 665nm peak of Ce6 after five minutes in the presence of rhodopsin.

3.3.2.3: Addition of DMSO

Adding Ce6 involves introducing small amounts of DMSO to the sample. Most of the samples had <1% DMSO added, so to control for any DMSO effects, DMSO was added to a final concentration of 1% (w/v) and the bleaching protocol followed. This rate was used as the 0 Ce6 rate when comparing with Ce6 samples for this reason.

3.3.2.4: Addition of chlorin

Using the ratio of A_{664} between spectra in a timecourse to judge the chlorin spectrum to subtract performed well, with the caveats described above. There is a small amount of deviation in the Q band around 665 nm, but the contributions in the rhodopsin relevant regions from 650-350nm are well controlled. A more time-consuming option here would have been to have decomposed the spectra into the individual chlorin and rhodopsin peaks, but significant optimisation would be required as the Soret peak, which overlaps with the (in some cases much) smaller 380 nm retinal peak is skewed. The gains in accuracy measuring the rhodopsin peak would have been marginal however.

The bleaching of Ce6 alone is not uniform, with some peaks bleaching more than others, as can be seen in Figure 3.1c. This can be seen in contrast with the linear response across wavelengths to dilution in Figure 3.1b. The differences caused by bleaching could account for the deviations in spectra after Ce6 subtraction, as the subtraction was necessarily performed using an unbleached spectrum. This is a minor effect and is unlikely to significantly affect the later analysis after decomposition of spectra. If some greater sensitivity was required, exploring the nature of the uneven bleaching might provide an alternative to peak-by-peak decomposition of the spectra.

There is a definite effect of relative chlorin concentration on the bleaching of rhodopsin illuminated with wavelengths beyond 630 nm. As shown in figure 3.4b, the trend is an increased rate of bleaching with greater Ce6 concentration, rising to a maximum of ~4 times greater than DMSO alone. The effect was not reliable, with great variability between samples at roughly the same concentration. Figure 3.4b shows all the rates collected, which have not been averaged as there was variance from the intended concentration of Ce6 added.

There is some evidence that the bleaching of chlorin is itself affected by mixing with rhodopsin. Figure 3.5c shows that a great difference in the relative concentration of chlorin to rhodopsin causes changes in the chlorin bleaching after five minutes, but the effect is variable and may not be observed after protocol improvements.

3.3.2.5: Dark noise with Ce6

The method for subtracting the Ce6 contribution to rhodopsin absorbance spectra in the dark and in the light developed above opened the door to previously inaccessible investigations of the interaction between rhodopsin and Ce6 by absorbance spectroscopy. It removes changes in the spectrum due to changes in Ce6 and allows the observation of rhodopsin in a strong Ce6 background. Because earlier studies had found that rhodopsin is more stable as measured by circular dichroism in the presence of Ce6 (Fernanda Balem et al., 2009) a stabilization of tertiary structure and retinal protein interactions was possible. On the other hand, cone opsins sensitive to red light are more prone to dark noise than rhodopsin and the other cone opsins (Luo et al., 2011), so another possibility was that dark noise of rhodopsin in the presence of Ce6 might be increased because its interaction leads to higher sensitivity to red light. Thus, it was possible that Ce6 might either stabilize or destabilize rhodopsin. To test for thermal activation, or dark noise, of rhodopsin, the bleaching effects of Ce6 in the dark were measured as for bleaching by red light. As shown in figure 3.4c, the dark noise at least does not increase upon addition of Ce6 at 37^{o}C. Similar

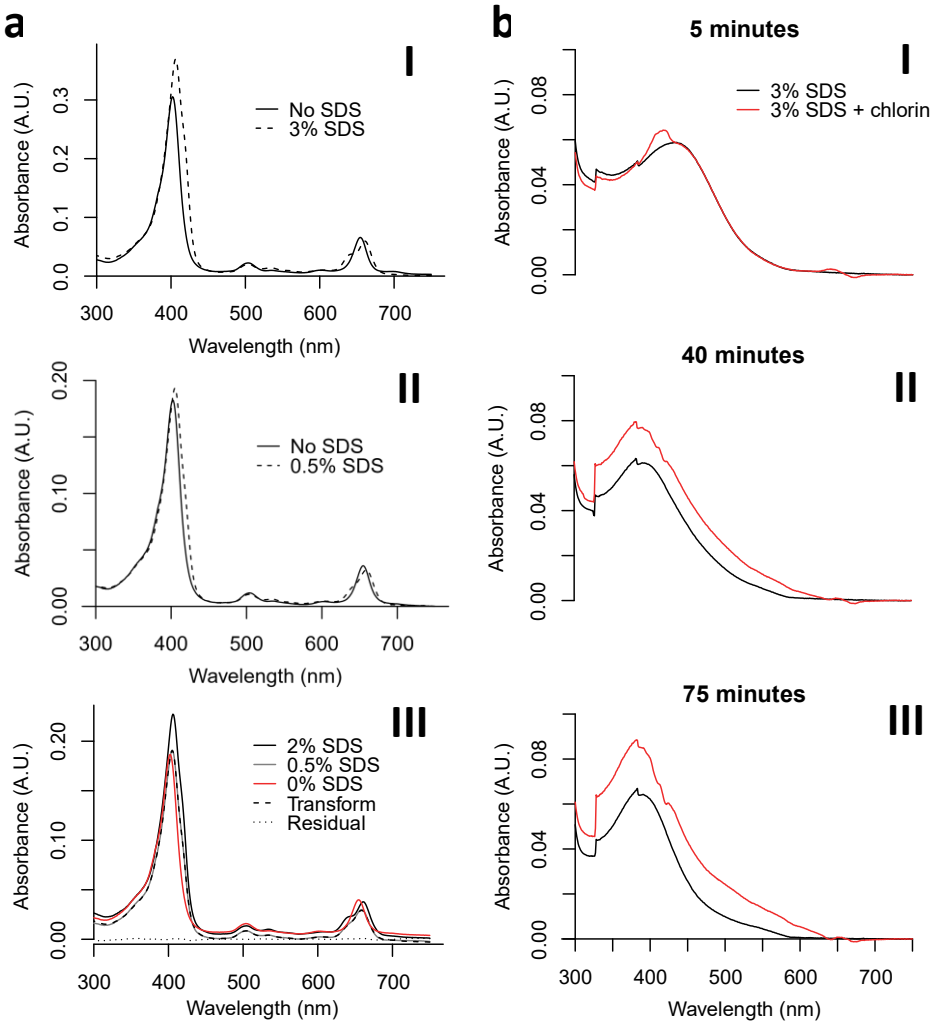


Figure 3.6: Chlorin and SDS a) Comparison of Ce6 spectra in the presence of **(I)** 3% and **(II)** 0.5% SDS (w/v) with SDS-free spectra. In **(II)**, measurements were taken before and after addition of SDS, and the 0.5% SDS spectrum is corrected for volume. In **(III)** the spectrum of Ce6 in 0.5% SDS was modelled as a combination of Ce6 spectra in 0 and 2% SDS. The Ce6 spectrum at 0.5% SDS was the sum of 0.62 times the same concentration of Ce6 at 2% SDS, and 0.30 times the same concentration of Ce6 at 0% SDS. **b)** Comparison of rhodopsin unfolding by 3% SDS in the presence of Ce6 at **(I)** 5 minutes, **(II)** 40 minutes, and **(III)** 75 minutes. Spectra with chlorin added (red line) have had the chlorin contribution subtracted as described for bleaching experiments.

results were observed at 25°C (data not shown). At 42°C, the loss of chromophore may be slightly slowed by addition of Ce6, though the difference is small and replication has not been attempted.

3.3.3: Chlorin and SDS

3.3.3.1: Chlorin SDS sensitivity

Another way of assessing protein stability is by quantifying its sensitivity to denaturation by SDS. Rhodopsin is often suspended in DM (Ramon et al., 2003), but can be denatured without aggregation by mixing with SDS (Dutta et al., 2010a). While assessing whether Ce6 affects the unfolding of rhodopsin by SDS, it was found that Ce6's absorbance spectrum is sensitive to SDS, which complicates analysis. The B-band is red-shifted by SDS, and the 665nm peak of

the Q-band is modified. With SDS, this peak seems to divide into two peaks, one around 640nm and another around 670. The effect seems to increase with more SDS (Figure 3.6a) up to 2% SDS but does not then change up to 3% SDS (data not shown). The intermediate concentration of 0.5% SDS can be modelled as a mixture of Ce6 spectra at 0 and 2% SDS. The relative intensity of the Q-bands to the B-band is also reduced, as can be seen in figure 3.6a I, where the major Q-band peaks are at around the same intensity, but the B-band peak is much higher with 3% SDS. When subtracting the contribution of Ce6 from rhodopsin experiments including SDS differences must be considered, and a fresh chlorin spectrum must be obtained, as previously acquired spectra will not be a good fit.

3.3.3.2: Chlorin and SDS with rhodopsin

The unfolding of rhodopsin by 3% SDS in the presence and absence of Ce6 was monitored by absorbance spectroscopy. The Ce6 contribution was subtracted from relevant spectra, and spectra compared. Figure 3.6b shows that there is little difference between spectra across time points between the two. There is some additional absorbance in the Ce6 spectrum, despite the rhodopsin concentration being the same in both samples, but as the λ_{max} is the same with and without chlorin (after subtraction of the Ce6 spectrum), it seems likely this is due to scattering.

3.3.4: ^{19}F NMR

One previously proposed hypothesis for the mechanism of red light sensitization of rhodopsin by Ce6 was through direct energy transfer (Ilyas Washington et al., 2004). Direct energy transfer from Ce6 to retinal would require proximity, and binding within ~10Å (Kvíčalová et al., 2016). Binding so close to the retinal pocket might be expected to affect the environment of retinal, and thus the NMR spectrum of rhodopsin reconstituted with a fluorinated derivative of retinal, so ^{19}F NMR spectroscopy was used to compare the environment of 14-fluoro-9-*cis*-retinal bound to rhodopsin in the presence and absence of Ce6.

3.3.4.1: Reconstitution

14-fluororetinal has previously been used to reconstitute rhodopsin (Colmenares, 1991; Steinberg et al., 1993). First, retinal was added in the dark prior to illumination, but the yield of reconstitution was negligible, as shown in figure 3.7a (I), which shows spectra of mixtures of rhodopsin and fluororetinal before and after illumination, including a spectrum taken after 12 hours. Adding fluororetinal after illumination raised yield to 25.1%, so further reconstitution was carried out in this way. Initially, fluororetinal was added so the change in

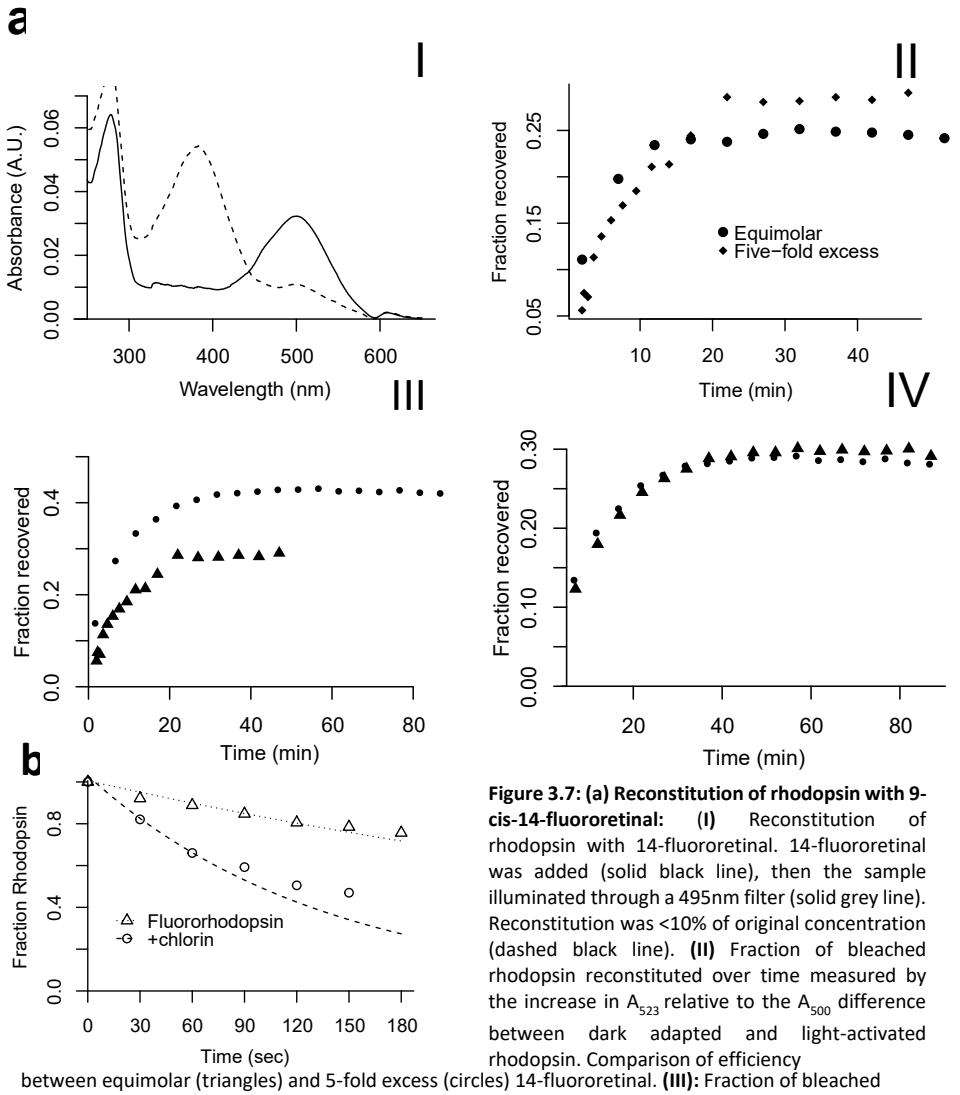


Figure 3.7: (a) Reconstitution of rhodopsin with 9-cis-14-fluororetinal: (**I)** Reconstitution of rhodopsin with 14-fluororetinal. 14-fluororetinal was added (solid black line), then the sample illuminated through a 495nm filter (solid grey line). Reconstitution was <10% of original concentration (dashed black line). **(II)** Fraction of bleached rhodopsin reconstituted over time measured by the increase in A_{523} relative to the A_{500} difference between dark adapted and light-activated rhodopsin. Comparison of efficiency between equimolar (triangles) and 5-fold excess (circles) 14-fluororetinal. **(III):** Fraction of bleached rhodopsin reconstituted over time measured by the increase in A_{523} relative to the A_{500} difference between dark adapted and light-activated rhodopsin. At 20°C (triangles) yield is 29.3%, at 4°C (circles), 42.5%. **(IV)** Comparison of A_{523} change between samples with (triangles) and without added Ce6 (circles). Addition of Ce6 and retinal at the same time complicates spectra too much to compare light activated spectrum, so time points are compared with the 2 minute measurement. **(b) Bleaching of fluororhodopsin with red light** monitored with absorbance spectroscopy as with rhodopsin. The wavelength measured was 523nm rather than 500nm, as the fluorinated chromophore has a maximum at this wavelength. The enhancement of bleaching rate is 3.92-fold

A_{380} was equal to the A_{500} of the original sample, with the aim of rhodopsin and fluororetinal being equimolar. When this failed to reconstitute most rhodopsin, excesses of fluororetinal were added. Adding 5X excess fluororetinal raises yield to 29.3% at 20°C (Figure 3.7a II).

Rhodopsin solubilised in DM, though equally stable in its dark-adapted state compared with native lipid, is less stable in Meta II, with an increase in Meta III formation (Ramon et al., 2003). To further increase yield, the temperature of reconstitution was reduced, and it was found that carrying out the reconstitution at 4°C increases yield to 42.5% (Figure 3.7a III). The decrease in temperature decreases the loss of reconstituting material to aggregation of

opsin (Lewis et al., 1997; Sudhoelter et al., 1982). Maximum yield was achieved within ~30 minutes.

Previous experiments reconstituting rhodopsin with 14-fluororetinal have had a reported efficiency of 70-100% (Steinberg et al., 1993). However, these reconstitutions were carried out on rod outer segment (ROS) suspension rather than detergent solubilised rhodopsin. It was also treated with NH_2OH, which rapidly reacts with the Schiff base to remove retinal, followed by extensive washing steps. The loss of yield may then have to be accepted as the price of reconstitution *in situ*.

Adding chlorin

Ce6 was added with fluororetinal during reconstitution at 4°C. As reconstitution efficiency is reduced the longer the time between illumination and addition of retinal, both Ce6 and fluororetinal were added at the same time, so the differences introduced by adding them are complex: a mix of Ce6, free retinal and whatever quantity of rhodopsin was reconstituted in the dead time. For this reason, the reconstitution with and without chlorin were compared from two minutes onwards (figure 3.7a IV) and decomposition was not attempted. The fractions reconstituted from this timepoint were very similar, so it was assumed that Ce6 has no effect on reconstitution.

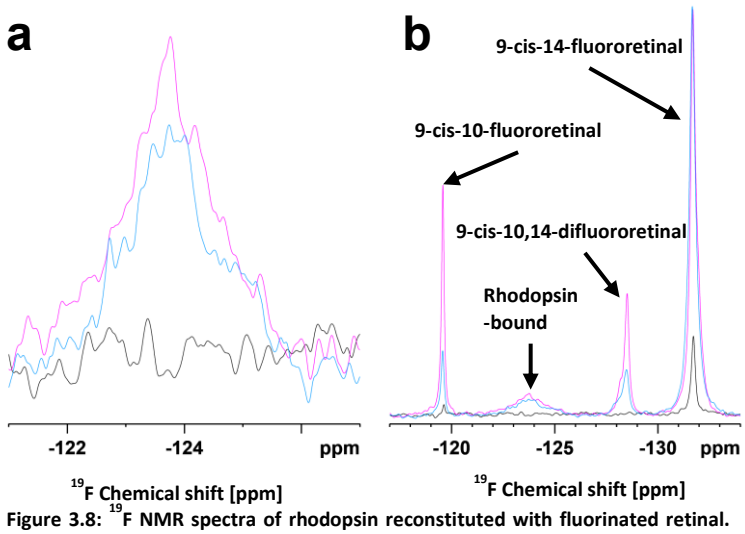


Figure 3.8: ^{19}F NMR spectra of rhodopsin reconstituted with fluorinated retinal. 1D 19F NMR of 150 µM [13C5, 15N]methionine 9-cis-14-fluororhodopsin in DDM, in the absence (magenta; 54318 scans; rg = 912) or presence (blue; 60110 scans; rg = 912) of chlorin e6, and 150 µM 9-cis-14-fluororetinal in 1.33% DDM in the absence of chlorin e6 (black; 50176 scans; rg = 256). All spectra were processed with line broadening of 80 Hz. The ^{19}F chemical shift of rhodopsin-bound 9-cis-14-fluororetinal is -123.76 ppm, with linewidths of 1.92 ppm and 1.75 ppm in the absence and presence of Ce6, respectively. Other peaks correspond to recorded values for free 9-*cis* retinal (Colmenares, 1991) with fluorines at positions 10 (-119.6 ppm), 10 and 14 (-128.5 ppm), and 14 (-131.7 ppm). **a)** shows the region containing the Rhodopsin-bound peak in the recorded spectra in **b)**

Chlorin's ability to photosensitize fluororhodopsin was tested as above, and 3.92 fold enhancement was measured with 29.4:1 Ce6:rhodopsin (Figure 3.7b).

3.3.4.2: NMR spectra

The ^{19}F NMR spectra of rhodopsin carrying the fluorinated retinal in the presence and absence of Ce6 are shown in Figure 3.8. There are four peaks present in these spectra (Figure 3.8b). The retinal peak at -123.76 ppm is essentially unaffected by Ce6 binding (Figure 3.8a), supporting the conclusion that Ce6 most likely does not bind near the retinal, and nor does it induce major structural changes that lead to perturbation of the retinal environment. The peak at -128.5ppm present in fluororhodopsin spectra is not found in the fluororetinal spectrum and is found at different intensities between spectra with and without chlorin. There are also differences in intensities between the peaks at -119.7 and -131.7. The -119.7, -128.5 and -131.7 peaks are similar to peaks reported for 9-*cis*-10-fluororetinal, 9-*cis*-10,14-difluororetinal and 9-*cis*-14-fluororetinal respectively (Leticia U. Colmenares et al., 1996). The apparent increase in retinal analogues other than 9-*cis*-fluororetinal in the absence of chlorin is surprising, as the same aliquot of fluororetinal was divided in two and added to each sample. Despite the apparent variation of the composition of free retinal in the sample, the broad peak of fluororetinal bound to rhodopsin is the same between samples, showing that any influence of Ce6 on rhodopsin does not affect the chemical environment of retinal.

3.4: Discussion

G protein conservation analysis. Prior work has suggested that binding of Ce6 to rhodopsin inhibits G protein activation (Gardner, Tirupula, Yanamala and Klein-Seetharaman, unpublished results). If the sequence of dragonfish rhodopsin had been significantly different in the region binding Ce6, it could be proposed that these differences could reconcile the fact that bovine rhodopsin's G-protein activation is compromised by Ce6 binding and an increase in signal in dragonfish. Analysis of a broad swathe of species' rhodopsins show that this region (as much of it is part of the G-protein binding site), is well conserved. Most of the residues are identical, though some are missing from the sequence, so this would benefit from a fuller sequencing of *M. niger*. This similarity shows that such an argument cannot be made. Other considerations are described below.

Red light sensitivity enhancement. The development of a subtraction method able to separate Ce6 and rhodopsin spectra has made higher ratios of Ce6 to rhodopsin accessible to experiment and has provided two conclusions. The first is to show that Ce6 enhances bleaching by red light in a dependent fashion. The variability between rate increases is so far

unexplained. The effect had also not saturated at the concentrations used. Ce6 binding to rhodopsin is in the micromolar range (around 8 μM, unpublished data), so much greater enhancement might be achievable by higher Ce6:rhodopsin ratios. Previous analysis did not separate Ce6 and rhodopsin bleaching, and has called Ce6 "unbleachable". This can be seen not to be true, as the spectra of Ce6 alone will show bleaching. Whether or not this is truly affected by the presence of rhodopsin is uncertain, as it does not vary strictly with rhodopsin:chlorin ratio. It is possible that, upon further examination, this effect may be simply dependent on the concentration of Ce6 itself, and perhaps the singlet oxygen produced by Ce6 excitation reacts with other Ce6 molecules, producing the observed bleaching. The rhodopsin bleaching rate enhancement can still be observed in the rhodopsin spectra with Ce6 subtracted. The subtraction method is not necessary to make this conclusion, but the concentrations of chlorin accessible to experiment are increased. However, the poor reproducibility of the experiment itself is a limiting factor here. The second conclusion is to show that rhodopsin's spectrum is not affected by Ce6 binding, as the bleaching, once the Ce6 contribution has been removed, proceeds as expected for rhodopsin: with a bleaching peak at 500nm, and an increasing peak at 380nm. That the retinal binding pocket is not directly affected by Ce6 is backed up by the lack of differences between ^{19}F NMR spectra of fluororhodopsin with and without Ce6.

If the *in vivo* sensitivity enhancement effects of Ce6 are moderate, it is possible that the stabilization of rhodopsin by Ce6 extends to reduction of dark noise, thus increasing signal:noise ratios. By lowering overall G-protein activation, it may yet increase red sensitivity through a combination of this effect and increasing the proportion of rhodopsin activations caused by red light. The adaptations of deep-sea fish eyes often show a pattern of sacrificing overall photon gathering capability in favour of greater discrimination as discussed in the introduction. Some dark noise reduction can tentatively be observed in samples with Ce6 added, though the data have not been replicated.

The additional stability conferred upon rhodopsin by Ce6 does not extend to SDS denaturation, which appears to proceed as normal. This data would need replication to draw a proper conclusion, but as this was the null hypothesis of the experiment, the line of enquiry was not pursued.

CHAPTER 4: ESTABLISHMENT OF PSRII AS AN EXPERIMENTAL MODEL SYSTEM FOR UNFOLDING

4.1: Summary

Previous theoretical considerations had indicated the unfolding of *Natronomonas pharaonis* sensory rhodopsin II (pSRII) may have similarities to both that of mammalian rhodopsin (MR) and bacteriorhodopsin (bR). The unfolding of pSRII by SDS was therefore monitored for the first time experimentally by UV/Vis absorbance spectroscopy, fluorescence emission at 335nm, circular dichroism and NMR spectroscopy. The circular dichroism data revealed a single first order process in the small loss of secondary structure. The absorbance and fluorescence data held information about the kinetics and physical characteristics of the tertiary structure changes which allowed the construction of a model of pSRII unfolding. Kinetics models were combined with existing knowledge of the interactions of retinal with retinal binding proteins to derive kinetic and photophysical parameters for this model across SDS concentrations.

Reversing this unfolding by dilution of SDS with 1,2-diheptanoyl-sn-glycero-3-phosphocholine (c7-DHPC) was observable by absorbance spectroscopy over minutes and hours, timescales amenable to similar analysis. The rate and extent of this refolding were useful for extending the model of unfolding to include reversible and irreversible unfolding.

Peaks of the NMR spectra recorded of pSRII unfolded in different concentrations of SDS display changes in chemical shift and peak intensity as a function of SDS concentration. The overall intensity of spectra decreases with higher SDS, so fewer peaks can be distinguished from the noise. Principal components analysis has been previously used to interpret NMR spectra and was applied here to the chemical shifts of assigned peaks. Physical interpretations of PCA distinguished two opposing patterns in the data with increasing SDS, and their implications for the folding mechanisms of pSRII are discussed. An additional intended purpose of this PCA was to add values for missing peaks in NMR spectra and predict where peaks going below the noise threshold at high SDS concentrations are expected. However, the low proportion of values remaining in the dataset made predictions inaccurate, so attempts to use image analysis techniques were made for this purpose with a dataset with limited resolution, showing that similar approaches may be useful for tracking signals across spectra in different conditions.

4.2: Introduction

4.2.1: Unfolding of MR by SDS

As discussed in section 1.1.2, MR folding is not adequately explained by the two-stage hypothesis. Ideally, the folding of MR would be explored with the refolding of the protein after unfolding, but the refolding of MR from unfolded states has not been possible. Examination of the unfolded states of proteins has been useful in the investigation of protein folding, as the remaining interactions in the residual structure of the unfolded state are indicative of interactions stabilizing the native structure in folded protein (Judith Klein-Seetharaman et al., 2002). For this reason, a range of denaturants were tested for their ability to unfold MR and keep it stable in unfolded states (Dutta et al., 2010b). Though sodium dodecyl sulphate (SDS) was not the most effective denaturant as determined by loss of secondary structure, unlike the other denaturants, the unfolded states did not aggregate, so SDS was used to characterize unfolded states of MR solubilised in DM (Dutta et al., 2010a). The unfolded states were then explored by absorbance spectroscopy, which showed the rapid loss of retinal-protein contacts at low concentrations of SDS, followed by the hydrolysis of the Schiff base covalently linking retinal to opsin. There was also loss of the quenching of the fluorescence of the five intrinsic tryptophans by retinal, indicating loosening of the helical bundle. These tryptophans are in close contact with retinal in the native structure. This data, with other biophysical techniques, indicated the formation of unfolded states with different characteristics dependent on the geometry of SDS micelles.

This work was followed up with NMR and electron paramagnetic resonance (EPR) spectroscopy of unfolded states of MR, demonstrating differences between the dynamics of the cytoplasmic domain when compared with the transmembrane and extracellular domains. For further discussion, see the introduction to chapter 5.

4.2.2: Least squares fitting

Physical data can be fitted to a mathematical model by the method of least squares, first used by the French mathematician Legendre (Merriman, 1877), and to track the asteroid Ceres as it passed behind the sun by the German mathematician Carl Friedrich Gauss (Stigler, 1981). A model for the data is formulated from knowledge of the system's behaviour, and least squares fitting chooses values for the model's parameters that minimise the squares of the residual error between the model and the data. For most models formed of a linear combination of parameters, a unique solution can be found, but for non-linear models, parameters are estimated by iterative refinement. The iterative nature of non-linear least squares (NLS) fitting means that the algorithms used will find a local minimum for residuals,

but not necessarily the global minimum. This becomes more important as the number of parameters increases. For the optimization of the parameters, initial estimates must be supplied, and as the likelihood of not finding the global minimum increases, the importance of good initial estimates increases. Poor estimates for complex models will result in finding a poor solution or not finding a minimum at all. Initial guesses very far from any minimum will not have any nearby estimates offering any improvement, and the fitting will not start.

The Gauss-Newton algorithm is commonly used for NLS problems, but this method is not robust against poor guesses. The Levenberg-Marquardt (LM) algorithm alternates between the Gauss-Newton algorithm and another method, the method of gradient descent, which makes it more robust (Marquardt, 1963). This makes it especially useful when attempting to automate fitting, as guesses supplied by algorithm may be poorer than those on a case-by-case basis.

There are many software programs capable of deconvoluting absorbance spectra by using NLS to fit the spectra to a model. For maximum control over fitting and for incorporation of kinetic models, scripts using the minpack.LM (Elzhov et al., 2016) package for R (R Core Team, 2015) were written for the purpose.

4.3: Methods development

The retinal chromophore makes a serendipitous tertiary structure reporter, but the overlapping absorbance peaks produced by the conformational states of unfolding pSRII and the single-dimensional nature of the fluorescence emissions measured over time pose challenges for analysis of the experimental data. However, non-linear least squares fitting, when combined with photophysical knowledge of the system for absorbance data and model selection methods for fluorescence data respectively allows the separation of consecutive reactions in the unfolding of pSRII, as described in detail in this chapter. For descriptions of methods used, refer to section 2.3.

4.3.1: Kinetics derivation

All SDS denaturation experiments were measured as a function of time. For a first order reaction $A \xrightarrow{k} B$, the rate law for [A] is that $\frac{-\delta[A]}{\delta t} = k[A]$. This rate law can be integrated to $\ln[A]_0 = \ln[A]_0 - kt$ and rearranged as $[A]_t = [A]_0 e^{-kt}$ to find [A] at a given time.

For two consecutive first order reactions, $A \xrightarrow{k_1} B \xrightarrow{k_2} C$, the rate law for A is unchanged, but the rate law for B depends on both rate constants (k_1 and k_2) and the concentration of A, such that $\frac{\delta[B]}{\delta t} = k_1[A] - k_2[B]$. C depends on the concentration of B and k_2, so the rate law

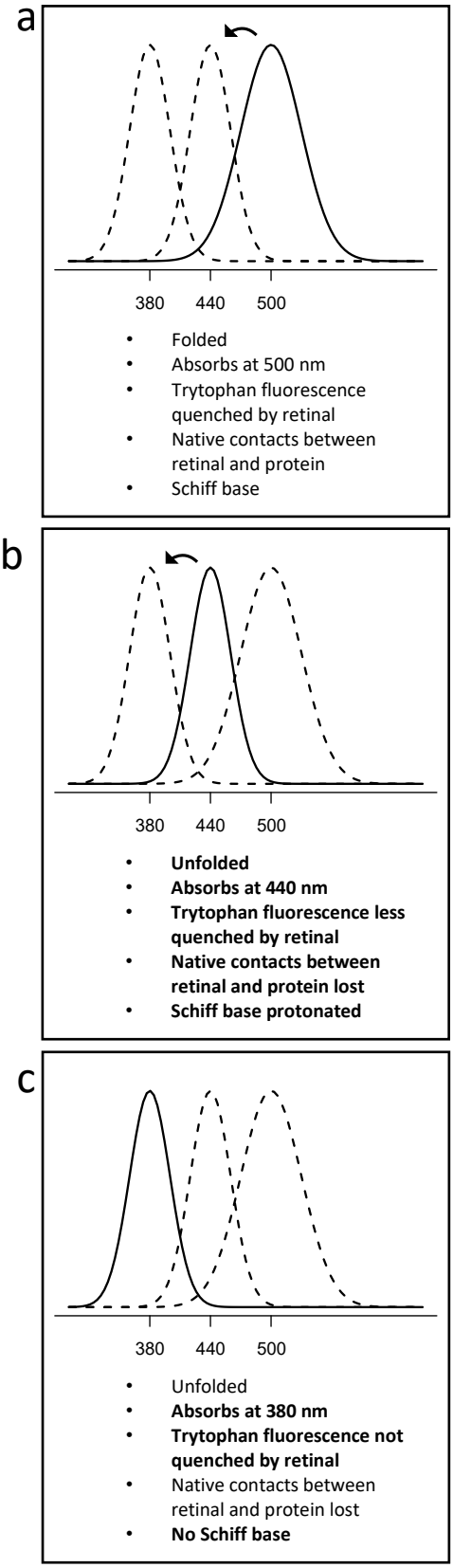


Figure 4.1: Schematic of the differences between unfolding species. (a) Species A, native pSRII. **(b)** Species B, unfolded pSRII with retinal attached. **(c)** Species C, unfolded pSRII and free retinal

is $\frac{\delta[C]}{\delta t} = k_2[B]$. The differentiation of these rate laws is more complicated but can be found in many physical chemistry textbooks (Atkins et al., 2014).

In this case, species A is the folded protein, in which retinal is covalently bound to the protein by a Schiff base linkage to a lysine, and native contacts between retinal and the protein are maintained. In species B, the specific native contacts have been lost, but the covalent linkage remains, though the Schiff base has been protonated. Species C represents free retinal after the Schiff base has been hydrolysed. The ways these species can be distinguished by both absorbance and fluorescence are discussed below, and summarised in Figure 4.1.

The derivation of these equations from the rate laws by differentiation means that the reactions can be modelled as:

$$[A]_t = [A]_0 e^{-k_1 t} \quad (1)$$

$$[B]_t = [A]_0 \left(\frac{k_1}{k_2 - k_1} \left(e^{-k_1 t} - e^{-k_2 t} \right) \right) \quad (2)$$

$$[C]_t = [A]_0 \left(1 + \frac{1}{k_1 - k_2} \left(k_2 e^{-k_1 t} - k_1 e^{-k_2 t} \right) \right) \quad (3)$$

where $[X]_t$ is the concentration of species X at time t. Under conditions when $k_1 << k_2$ the contribution from [B] will be negligible, as the $\frac{k_1}{k_2 - k_1}$ term will become very small. Concentrating on the contribution from [C],

$$[C]_t = [A]_0 \left(1 + \left(\frac{1}{k_1 - k_2} \right) \left(k_2 e^{-k_1 t} - k_1 e^{-k_2 t} \right) \right)$$

The $-k_1 e^{-k_2 t}$ term will be much smaller than $-k_2 e^{-k_1 t}$, as k_2 will be large enough that $e^{-k_2 t}$ will be close to zero.

$$[C]_t = [A]_0 \left(1 + \left(\frac{1}{k_1 - k_2} \right) \left(k_2 e^{-k_1 t} \right) \right)$$

In $\frac{1}{k_1 - k_2}$, k_1 will be negligible, and k_2 will cancel in $k_2 e^{-k_1 t}$, leaving

$$[C]_t = [A]_0 \left(1 - e^{-k_1 t} \right) \quad (4)$$

In the context of the reactions under examination here, this means that when the rate at which retinal loses specific contacts with the protein (k_1), forming the protonated Schiff base absorbing at 440nm, is lower than the rate at which the Schiff base is hydrolysed (k_2), forming free retinal which absorbs at 380nm, the 440nm species might not be able to accumulate sufficiently to be observed.

4.3.2: Absorbance

Most absorbance spectra can be modelled as a combination of Gaussian peaks of the form

$$Abs_\lambda = \sum a_x e^{-\frac{(\lambda - c_x)^2}{2w_x^2}} \quad (5)$$

Where a_x is the amplitude, or absorbance at the peak's centre c_x, equivalent to the λ_{max}. The peak's width is defined by its half width at the half maximum, w_x. The absorbance peak of free retinal has a centre at 387 nm, very close to the deprotonated Schiff base absorbance maximum of 380 nm (Doukas et al., 1978). A protonated Schiff base has an absorbance maximum at 440 nm. The further red shift of the absorbance maximum in both folded pSRII

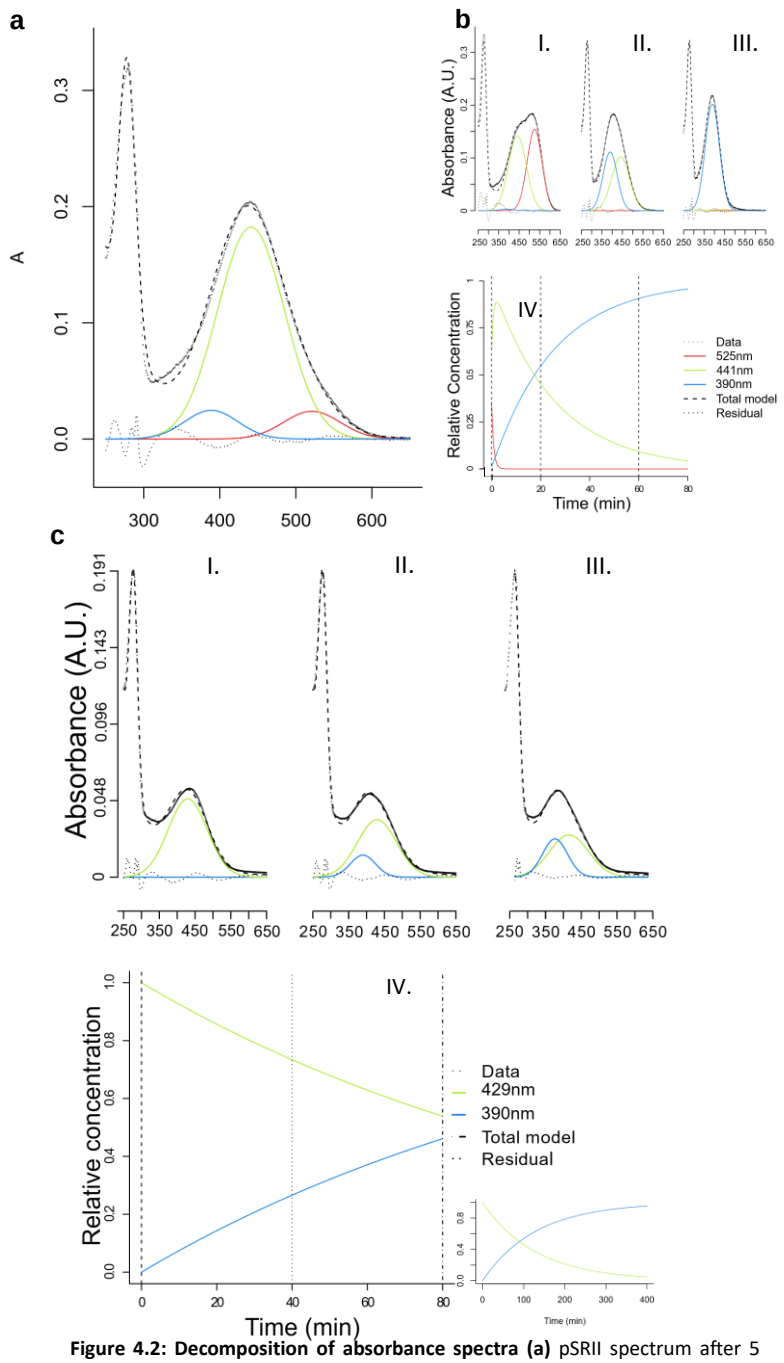


Figure 4.2: Decomposition of absorbance spectra (a) pSRII spectrum after 5 minutes unfolding in 1% SDS (dark grey dots), showing the native pSRII peak (red), the protonated Schiff base peak (green) and the free retinal peak (blue). The dashed line shows the total model, including a 280nm peak and a scattering baseline (not shown). The residual of the model is shown (dotted line). **(b)** Example timecourse of pSRII unfolding in 3% SDS. **(I), (II),** and **(III)** show 0, 20 and 60 minutes respectively. Spectra are displayed as in (a). **(c)** Example timecourse of MR unfolding in 3% SDS for comparison. **(I), (II),** and **(III)** show 0, 40 and 80 minutes respectively. The reaction was fitted as a single first order reaction. Inset shows the progress of the reaction from 0 to 400 minutes.

(to 498nm) and MR (to 500nm) is due to further spectral tuning by the protein (Bravaya et al., 2007; Ren et al., 2001; Teller et al., 2003). In the unfolding of these proteins, a mixture of ~500 nm, 440 nm, and 380 nm peaks can be observed due to the presence of retinal in native

protein, retinal which has lost specific protein contacts and retinal hydrolysed from the protein.

The relative contributions of the different peaks across time can be measured by recording many timepoints during the unfolding reaction. By applying the Beer-Lambert equation to the peaks, the amplitude of the peak can be converted into concentrations so, if a_x is the absorbance of a species at a given wavelength, [X] is the concentration of that species, and ε_x is its extinction coefficient at that wavelength, $a_x=[X]\varepsilon_x$. Assuming the three peaks represent all the retinal present in a cuvette and that the overall concentration does not change over time, application of mass action law means

$$[Retinal] = \frac{a_{380}}{\varepsilon_{380}} + \frac{a_{440}}{\varepsilon_{440}} + \frac{a_{500}}{\varepsilon_{500}} \quad (6)$$

The concentration and extinction coefficient terms can be substituted for the amplitude of the peaks in the decomposition of spectra. One approach for determining the kinetics is to decompose the spectra, then to fit the three amplitudes at each timepoint to the overall concentration as in the mass action law above. However, this is limited because the choice must be made either to fix the peak parameters and accept a potentially poorer fitting at each timepoint, or to allow the peaks to vary and risk a poorer fitting of the extinction coefficients. However, if a model for the kinetics of the reaction can be assumed, as in the case of the unfolding of retinal binding proteins, the kinetic equations for the reaction $Native \overset{k_1}{\rightarrow} Protonated\ Schiff\ base \overset{k_2}{\rightarrow} Free\ retinal$ (equations 1-3) can be combined with the decomposition of peaks, using equation 5, with the inclusion of extinction coefficients to convert to concentrations, using equation 6. An error term to handle the dead time when loading samples and existing absorbance around 380 nm can be included to give:

$$\begin{aligned} A \sim & \varepsilon_{500}\left([Total]e^{-k_1(t+f)}\right)e^{-\frac{(\lambda-c_1)^2}{2w_1^2}} \\ & + \varepsilon_{440}\left([Total]\left(\frac{k_1}{k_2-k_1}\right)\left(e^{-k_1(t+f)} - e^{-k_2(t+f)}\right)e^{-\frac{(\lambda-c_2)^2}{2w_2^2}}\right. \\ & + \varepsilon_{360}\left([Total](1 + \frac{1}{k_1-k_2}\left(k_2e^{-k_1(t+f)} - k_1e^{-k_2(t+f)}\right)e^{-\frac{(\lambda-c_3)^2}{2w_3^2}}\right. \end{aligned} \quad (7)$$

[Total] refers to the total retinal concentration, and f to the dead time of the experiment. Other symbols used here are from equations 1-6 above.

Fitting time courses of absorbance spectra to this equation allows optimisation of all parameters with the largest dataset. With the power of a modern desktop computer, the

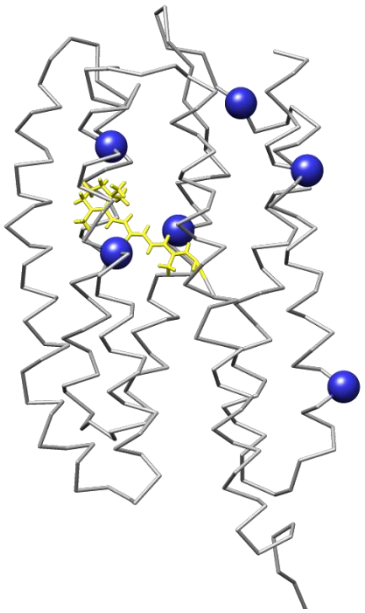


Figure 4.3: Structure of pSRII showing tryptophans as blue spheres and retinal in yellow sticks.

time taken to optimise the larger number of parameters is negligible. Scripts in R were written to automate the fitting, plotting, and extraction of parameters for unfolding time courses (as described in section 2.3.2.1). An example for pSRII unfolding is shown in Figure 4.2a.

4.3.3: Fluorescence

pSRII contains 6 intrinsic tryptophans (Figure 4.3), which were used to investigate its tertiary structure. Retinal quenches this fluorescence by a combination of Förster resonance energy transfer, which depends on distance between donor and acceptor groups, and dynamic quenching, which relies on physical contact between donor and acceptor (Alexiev and Farrens, 2014). Three of the six tryptophans in pSRII are within 1 nm of retinal, and close enough for dynamic quenching. This makes fluorescence emission at 335 nm upon excitation at 295 nm a reporter for the folding state of pSRII, as the distances between the tryptophans and retinal change during unfolding.

Exponential growth decay curves describe data which follow the same pattern as exponential decay, but increasing to some asymptote, rather than decreasing and are of the form:

$$y = A\left(1 - e^{-kx}\right) \tag{8}$$

where A describes the asymptote towards which the curve tends, henceforth called the amplitude of that curve. Initial parameters for A and k can be found from simple features of the data. Starting values for A can be found by simply taking the maximum value of y (y_{max}). This is more accurate if the data has been allowed to plateau.

The rate constant of the curve, k, can be expressed as $\frac{1}{\tau}$, where τ, the time constant, is related to the half-life of the curve as $\tau = x_{\frac{1}{2}}/\ln(2)$, so the time at which $y = \frac{A_{est}}{2}$ gives an estimate of τ. An exponential growth decay curve is then fitted to the data with the given starting values.

Fitting of an exponential growth decay curve to the fluorescence did not always give a good fit, as can be seen in Figure 4.4a. These time courses had a characteristic residual profile (shown as a green line in figure 4.4a), with an early positive region then a smaller negative

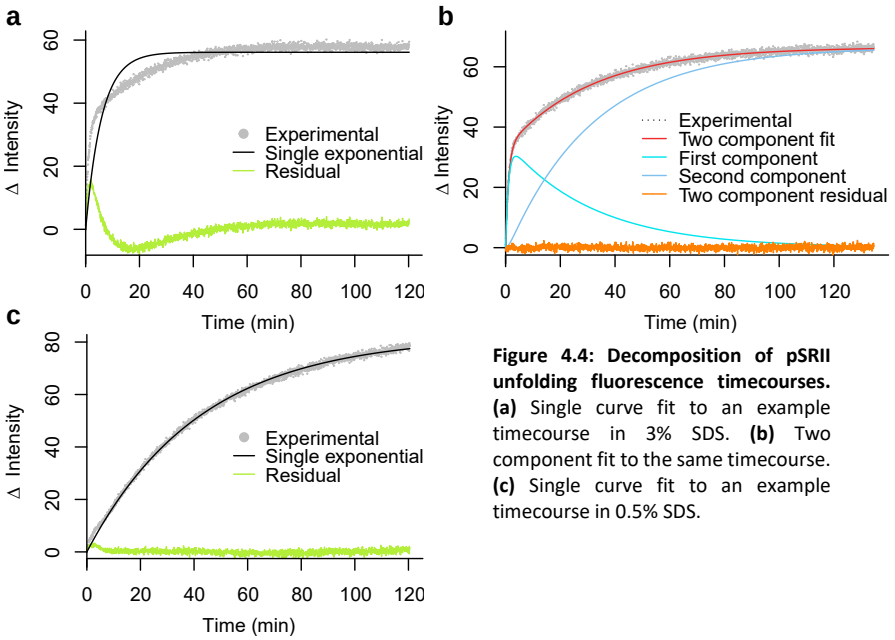


Figure 4.4: Decomposition of pSRII unfolding fluorescence timecourses. **(a)** Single curve fit to an example timecourse in 3% SDS. **(b)** Two component fit to the same timecourse. **(c)** Single curve fit to an example timecourse in 0.5% SDS.

region which diminishes so the residual at the plateau is around 0. Using the same estimation process as for the original curve on the residual yields a second, smaller exponential curve. Within the unfolding model derived from absorbance spectroscopy there is no physical explanation for two parallel processes. In addition, such a model would have to explain the division of single and double exponential processes by SDS concentration.

Referring to the kinetic equations above (equations 1-3), the concentrations of species B and C (protonated Schiff base and hydrolysed Schiff base) can each be seen as the superposition of two exponential curves, as each contains some multiplication of $e^{-k_1 t}$ and $e^{-k_2 t}$. Derivation of this can be found in section 2.3.2.1. This means that the two exponentials detected by fitting can be used as an approximation of the rate equations for B and C in consecutive first order reactions.

If each component in the consecutive reactions is assumed to have a different amount of quenching, then $Intensity = F_{Native}[Native] + F_{Protonated}[Protonated] + F_{Apoprotein}[Apoprotein]$. Here, intensity only increases, so we can say that $F_{Native} < F_{Protonated} < F_{Apoprotein}$. To simplify, we can instead look at $\Delta Intensity = I_{Protonated}[Protonated] + I_{Apoprotein}[Apoprotein]$, where $I_x = F_x - F_{Native}$, so I_{Native} = 0. This allows the contribution from the native protein to be ignored, as well as any change in the other reactants in the dead time of the experiment. Substituting in the kinetic equations for [Protonated] (equation 2) and [Apoprotein] (equation 3) yields the model

$$\Delta Intensity = I_{Prot}[Total]\left(\frac{k_1}{k_2 - k_1}\right)\left(e^{-k_1 t} - e^{-k_2 t}\right)$$
$$+I_{Apo}[Total]\left(1 + \left(\frac{1}{k_1 - k_2}\right)\left(k_2 e^{-k_1 t} - k_1 e^{-k_2 t}\right)\right) \quad (9)$$

To allow for the noise in the data, when fitting to ΔIntensity a constant value was added to the models so a fitted curve could start at values other than zero. The parameters from the biexponential growth model were adapted to provide starting values for this model. I_{Prot} ~A_2/[Total] and I_{Apo}~A_1/[Total]. The rate constants were close enough to use as initial estimates without any transformation.

Using the kinetic model to explain the appearance of two exponential curves also explains why the second curve is absent in some time courses. In section 4.3.1 above there is an explanation of why a low k_1 would make the apparent rate $[A]_0(1 - e^{-k_1 t})$, and as fluorescence only increases, this is what is observed at low SDS.

4.3.3.1: Model selection

When comparing a model with two curves to a model with one, the model with more parameters will always fit at least as well as the simpler model (i.e. amplitude could be set to 0 and the second curve makes no contribution), and is likely to outperform it. Even if the simpler model is closer to the truth, adding parameters will make it possible to fit to the noise of the data. A better measure than simply comparing the residuals is required. There are two types of method used to compare models.

4.3.3.1.1: Information criteria

The first approach is to impose a penalty on adding parameters to the model, and establish a criterion that has a term for the residuals and a term for number of parameters. Several model selection criteria exist based on this paradigm with the terms based on different theoretical considerations. Akaike's An Information Criterion (AIC) is the earliest and a popular model selection criterion (Akaike, 1974). It has its basis in minimising information entropy which is theoretically intensive, but thankfully simple to implement for least squares fitting:

$$AIC = 2K - nln(\hat{\sigma}^2)$$

Where K is the number of estimated parameters. This is equal to the number of model parameters plus one, as the variance of the data from the model is also estimated. $\hat{\sigma}^2$ is the estimated variance, which in this case is equal to the average of the residuals squared and n is the sample size. The AIC estimates the relative distance between candidate models and

the truth. Thus, a lower AIC is preferred, meaning the addition of a new parameter is penalised unless this refinement reduces the estimated variance enough to compensate for the increase in the parameter term. A shortcoming of using the AIC is that as sample sizes increase, a simple model eventually becomes disfavoured as small effects are magnified. A generalisation of information criteria exists that makes fewer assumptions than the AIC, but this loses the advantage the AIC has of being very easy to implement. When selecting models for the fluorescence data, the initial solution attempted was to take random samples of 500 points and perform model selection on these. Quite apart from being theoretically dubious, this solution also lost some of the simplicity of using the AIC and became closer to the other family of model selection methods, so was replaced with cross-validation.

The same analysis was carried out on molar residue ellipticity (MRE) data acquired from circular dichroism (CD). The AIC was used to compare fits between models as for fluorescence. As model selection criteria discriminate between models based on bias and variance (Hastie et al., 2009), if the variance for both models is very high, bias will become the discriminant and the more complex model will be rejected. As the noise was very high, the variance for both models was sufficiently high that it is likely even if two exponentials underlie the CD data AIC selected the simpler model every time.

4.3.3.1.2: Cross-validation

Cross-validation is a model selection technique often used in machine learning and data mining. The basic idea is to check for overfitting by fitting the model to some subset of the data (the training set), then testing how good the fit is on the rest of the data (the testing set). Any noise in the training set that the model fits to will introduce errors when compared to the testing set. Different versions of cross-validation with different methods for generating the training and testing sets exist, but for the fluorescence data a relatively simple method called k-fold cross-validation was used. Briefly, this splits the dataset into k equal sized sets (folds). One fold is removed from the data to make the training set and then used as the testing set. This is repeated for each fold, and the average residual sum of squares (RSS) for each candidate model is compared and the model with the lowest average RSS is selected. Details of the procedure can be found in the data analysis section of “Materials and Methods”.

Initially the AIC was used for model selection. After switching to 10-fold cross-validation, the model selection was compared for a subset of 9 of the previous results. Both methods selected the same model in each case, so no further testing took place. Examples of the application of this method to a timecourse for which the more complex model was more

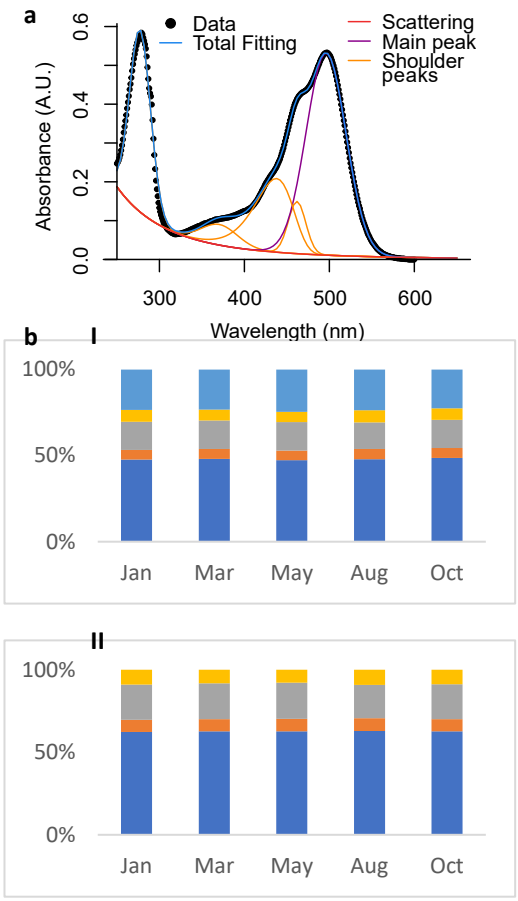


Figure 4.5: Fitting native pSRII spectra. (a) Example fitting of a native pSRII spectrum used for construction of a native profile. **(b)** Areas of the peaks in five native pSRII spectra as a percentage of the total area **(I)** Showing all peaks. The light blue bar shows the protein peak at 280nm. The yellow, grey, orange and dark blue bars show the chromophore peaks at 370, 420, 460 and 498nm respectively. **(II)** as (I), but omitting the contribution of the protein peak.

appropriate (a and b) and the less complex model was more appropriate (c) are shown in figure 4.4.

4.3.4: Refolding and hydroxylamine fitting

When fitting the unfolding absorbance data, the three peaks described above sufficed as a model. However, rather than the single peak at 498 nm used to model native pSRII above, spectra of truly native pSRII comprise four peaks: the main peak at 498 nm and three minor peaks at 460, 420 and 380 nm, which arise from vibronic effects in the system (Ren et al., 2001). These peaks disappear with concentrations of SDS above 0.5%. However, upon dilution of SDS, these peaks reappear in refolding pSRII. These make a non-negligible contribution to spectra and distort fitting if ignored. It was also unknown whether the contributions of these peaks vary across the refolding process. To determine refolding progress, therefore, it was necessary to construct a native profile of these peaks against which to compare. To this end, five native pSRII spectra had five Gaussian peaks fitted, including one for the 280 nm protein peak. An example is shown in Figure 4.5a, and the areas of each peak in the fitting are shown in Figure 4.5b and c. The inclusion of the 280 nm peak necessitated the fitting of the UV region with wavelengths below 350nm, previously excluded from analysis. In this region, scattering becomes significant, so a scattering baseline was added to the model. As scattering is proportional to λ^{-4}, the terms

$$D + \frac{E}{\lambda^G} \tag{10}$$

were added, which includes a linear term because scattering occurs to some extent at all wavelengths, so a zero value in the spectrum must be compensated for. Due to the vibronic nature of the additional retinal peaks, they are not symmetrical (Barker and Fox, 1980; Siano and Metzler, 1969). There are several lineshapes usable for asymmetric peaks, but the

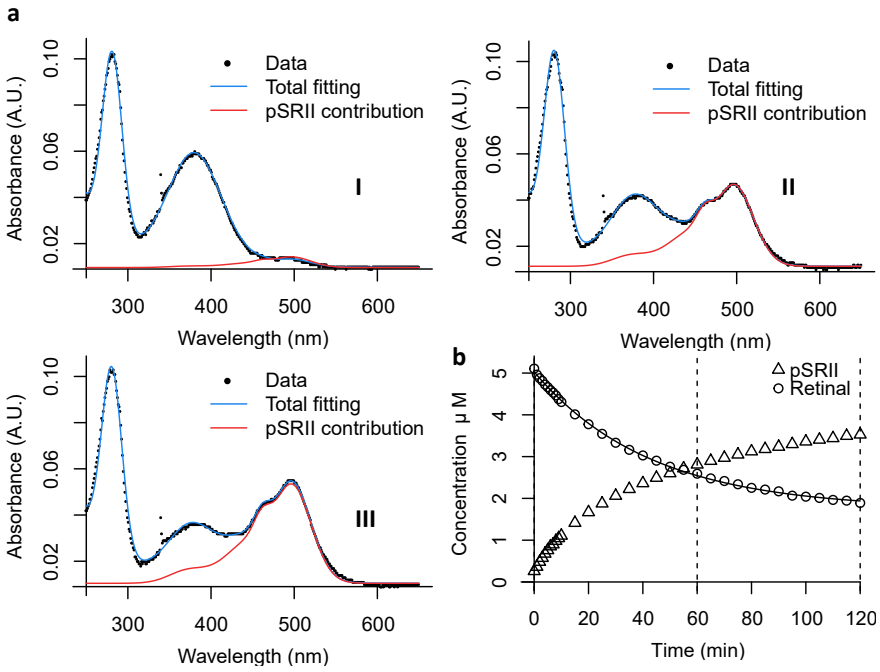


Figure 4.6: Fitting of refolding time courses. (a) Example fitting of refolding of pSRII by diluting SDS at **(I)** 0, **(II)** 60, and **(III)** 120 minutes. **(b)** Example kinetics of refolding pSRII, with the spectra in (a) indicated by dashed lines.

simplest implementation for NLS is to replace the width term with a sigmoidal function (Stancik and Brauns, 2008). This adds one parameter to the model which controls the skew of the peak, so a Gaussian peak like equation 5 becomes:

$$Abs_\lambda = \sum a_x e^{-\frac{(\lambda - c_x)^2}{2\frac{w_x}{1+e^{r(\lambda - c_x)}}}} \qquad (11)$$

All terms from equation 5 are the same, with the addition of the parameter r, which controls the skew.

Due to the skew of the peak, the amplitude parameter is no longer directly comparable between peaks to compare the contribution to the spectrum. To allow comparison between peaks between spectra, the peaks were instead integrated to find the area under the curve within the range of the recorded spectra. The areas of the peaks were calculated for the five native spectra. In each spectrum the areas were practically identical (figure 4.5b and c), and similar to published values (Ren et al., 2001).

As the kinetics of the different peaks were unknown, no combination of peak amplitudes and kinetic profiles was made for overall fitting so that no explicit kinetic model had to be assumed to carry out fitting. Instead, each timepoint was fitted separately, and later grouped by their centres. The areas of each peak were found and plotted over time. To determine whether the increase of the shoulder peaks is at a different rate compared to the 498 nm peak, the relative areas of the 380 nm peak and 498 nm peak were plotted against one

another and linear regression was carried out between the two (figure 4.12b). Though different experiments have different degrees of unfolding, causing them to cluster at different positions on the line, there was little deviation from the linear regression, showing that the main and the shoulder peaks increase at the same rate.

Because of the simple relation between the recovery of the native peaks and the decrease in the 380 nm peak, the progress of refolding was defined as

$$Refolded = 1 - \left(\frac{Ret_T - Ret_{native}}{1 - Ret_{native}}\right) \quad (12)$$

The kinetics were modelled as a first-order reaction. The recovery efficiency was found by finding the maximum recovery compared to the native profile (figure 4.12a).

Automation of fitting

When further refolding experiments were carried out, an automated approach was used. The earlier fittings showed that refolding spectra were combinations of a free retinal peak with a native-like profile of retinal peaks. For automation, an average fitted profile was established by fitting five native spectra, then adjusting the amplitudes of the peaks so that they sum to one, then taking the mean of each parameter. A model was created so that each spectrum in a time series was fitted as:

$$Absorbance \sim D + \frac{E}{x^G} + Peak_{A_{280}} + Peak_{A_{380}} + B \times Profile_{native} \quad (13)$$

Where D, E and G refer to the scattering terms in equation 10, the peaks at 280 and 380nm are as shown in equation 5, and the native profile is a mixture of three peaks as equation 5, and one as equation 11.

From these fittings, the amplitude of the 380 nm peak and B (the multiplier for the native profile) were taken to represent the concentrations of free retinal and refolded pSRII respectively. The overall concentration remains constant, and as the effective extinction coefficient of the native profile is not certain, the concentration of each species at each timepoint were estimated using non-linear least squares fitting by fitting those parameters to the model

$$Concentration \sim Ret \times A_{380} + pSRII \times B \quad (14)$$

The refolding was incomplete (see section 4.4.1.9), as indicated by absorbance at 380 nm beyond the native profile and recorded by residual amplitude of the fitted 380 nm peak (e.g

figure 4.6a). The concentrations determined from this amplitude were fitted to an exponential decay curve of the form

$$C \sim Ae^{-kt} + c \tag{15}$$

the *c* term of which allows for the fitting of incomplete decays. The residual concentration estimated by *c* was divided by the given overall concentration to estimate the fraction of pSRII incapable of refolding for that time course. Scripts for the automation of this fitting and plotting of the fits were written, tested on a sample dataset as shown in Figure 4.6, and sent to Yi Lei Tan at the University of Cambridge.

Unfolding of pSRII by SDS was also carried out in the presence of hydroxylamine, which reacts with exposed Schiff base to rapidly form a retinal oxime. The spectra of this reaction, unlike those in the absence of hydroxylamine, have no protonated Schiff base peak and maintain the shoulder peaks of native pSRII, so modelling the reaction is similar to modelling refolding pSRII, just in the opposite direction. The fitting procedure for native spectra was also used for pSRII in low concentrations of SDS.

4.3.5: Automatic peak picking

When originally fitting peaks to the refolding spectra, the parameters of the shoulder peaks were not immediately obvious and obscured by the much larger main peak dominating that area of the spectra. In case these parameters changed over time, an automatic peak picking strategy to provide initial estimates was pursued. The algorithm was designed to have the option of fitting a baseline, then to fit peaks until model selection found adding more to be extraneous. Ultimately, the script was not used for refolding spectra because of the skew of the vibronic peaks. No skewing parameter was originally implemented, and after a literature search identified the source and nature of the shoulder peaks, the script was unnecessary. However, when fitting only symmetrical peaks, the algorithm works, and by using clustering strategies, changes over time can be estimated. In future, it is possible that a refinement of the original strategy could be extended to automatically fit kinetics to time courses of absorbance spectra.

4.3.5.1: Estimation strategy

A full description of the peak picking strategy can be found in section 2.3.3. What follows is a brief summary. Data can be smoothed by calculating the average of the values around that point. This can particularly distort peaks, so was used, in combination with a geometrical filter (figure 4.7b), to identify candidate peak positions. A second step used NLS fitting for each of these candidates to refine the estimates for a peak in that position. The final step is

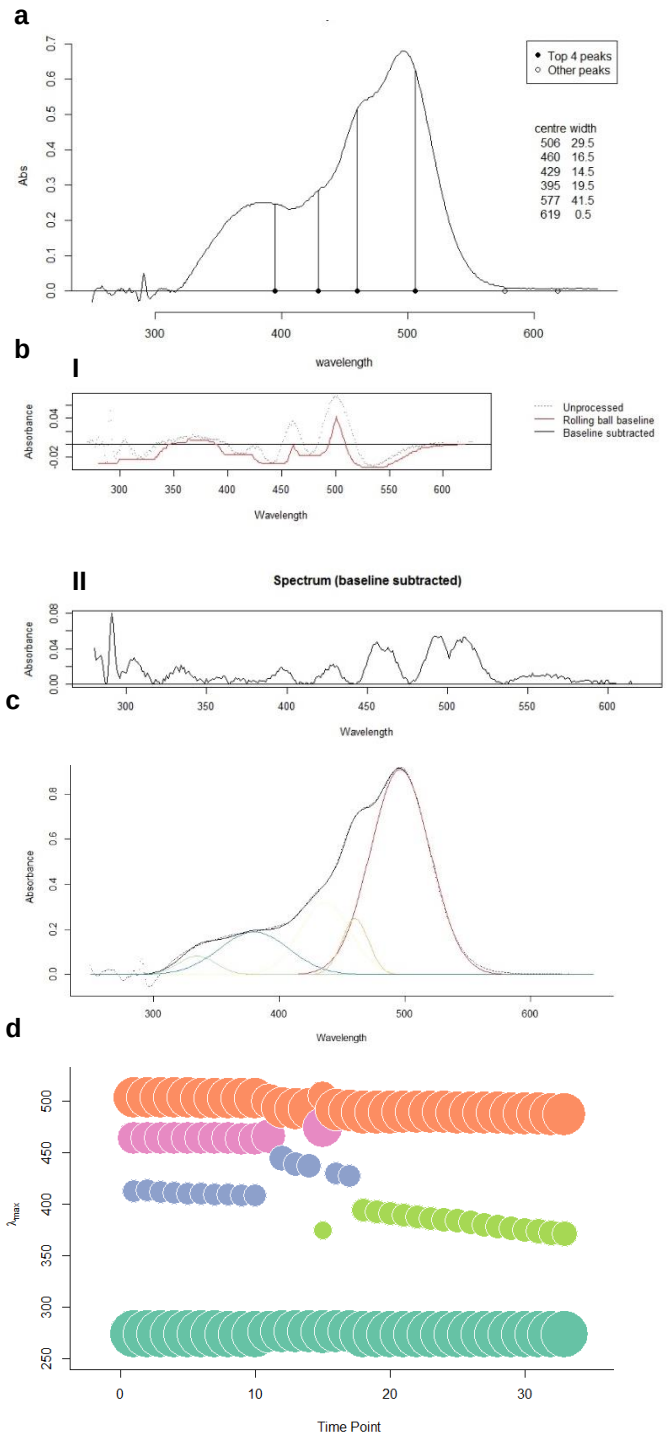


Figure 4.7: Automatic peak picking. (a) Peak centres acquired from oversmoothing algorithm, indicated on a native pSRII spectrum. The protein peak at 280nm, due to its separation, was easier to fit and constant across time points. **(b)** Intermediate steps of oversmoothing: **(I)** Difference between smoothed spectrum and raw spectrum, showing the baseline identified by the rolling ball filter. **(II)** The difference spectrum after baseline subtraction. **(c)** Native pSRII spectrum with peaks from nls fitting using oversmoothed guesses. **(d)** Peak parameters for a refolding timecourse. Areas of circles indicate the area of the peak, with y-axis position indicating the peak centre. Circles are coloured according to hierarchical clustering.

to establish the number of peaks present by adding these peaks to a total model one by one, comparing the AIC between models with and without each peak to establish whether it is

necessary. Using this, spectra can be automatically fitted very quickly: 100 spectra are fitted in an average of 4.5 seconds on a desktop computer, even with the difficulties caused by skewed peaks (figure 4.7c).

4.3.5.2: Clustering

Treating the final fitted peaks as points in 4-dimensional space, with amplitude, centre, width and time as dimensions, peaks across timepoints can be grouped into clusters. In this way, the progress of different species might be measured. In the current implementation, a number of clusters must be defined manually. There are more advanced clustering methods that may be able to detect the number of clusters required. Once the peaks have been clustered, average parameters for each peak across timepoints could be used as refined estimates for each spectrum with the aim of eliminating outliers and measuring timepoints more reliably. On the test data, the algorithm did not always fit realistic peaks, with some spectra being dominated by one enormous peak. It is possible these might disappear with an improved initial estimation but given the large set of similar spectra it seems wasteful not to use the aggregate behaviour to optimise peak picking. Initial attempts were not accurate in their clustering, as can be seen in Figure 4.7d, though this may have been due to the poor fitting from the lack of a skew parameter in the underlying model.

4.3.5.3: Fitting skewed Gaussians as symmetrical

From the fitting carried out, it can be seen that a skewed peak can be approximated by symmetrical Gaussians. The parameters of two peaks fitted to a simulated skewed peak vary in regular patterns which might be useful for recognising skewed peaks in real data. However, the ways a skewed peak and other, nearby peaks combine may make fitting less predictable, as can be seen by the skewing of the 460 nm shoulder peak in pSRII causing the 380 nm peak to be split into two while remaining a single peak itself, as can be seen in Figure 4.7c.

4.4: Results

4.4.1: Biophysical kinetic studies of pSRII

4.4.1.1: Early changes in absorbance spectra

pSRII samples took around a minute to load into the spectrophotometer. Within this time, especially at higher concentrations of SDS, some changes in the spectra had occurred. The transition from a native spectrum to a protonated Schiff base spectrum had already started, but there were other changes.

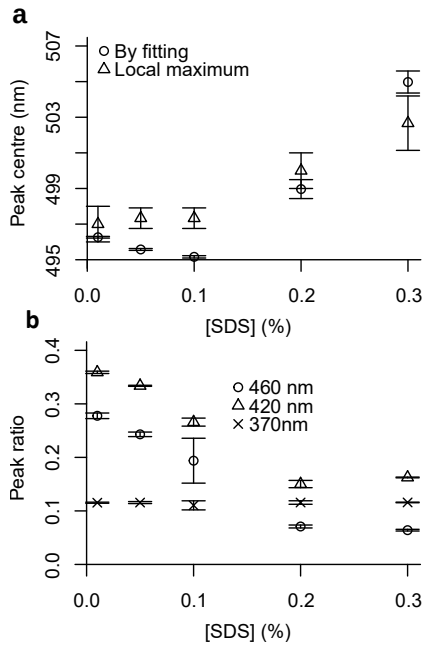


Figure 4.8: Early spectral changes of pSRII in SDS. (a) Changes in the centre of the main pSRII peak as measured by fitting of a five-peak model (circles) and by taking the local maximum (triangles). Data shown as the mean of three values +/- standard deviation. **(b) Changes in the balance of shoulder peaks** in the five-peak model. The ratio is of the area of that peak to the total area of all retinal peaks in the model. Data shown as the mean of three values +/- standard deviation.

First, the absorbance maximum of the native peak is red shifted, as shown in figure 4.8a. The spectra of pSRII in low concentrations were fitted in the same way as native spectra (section 3.4) The fact that with higher rates this peak is smaller in the first spectrum might cause some error in the fitting, but even at 0.5% SDS, where this peak is still large, the centre has shifted from 498nm to ~515nm. If the shift was a fitting artefact, the 440nm peak might be expected to red shift or widen to compensate, which does not happen. Figure 4.10c shows that beyond 3% SDS the peak centre is moved as far as 528nm.

The shoulder peaks present in native pSRII (indicated in Figure 4.5a) are also absent in spectra above 1% SDS. It is possible that they are just rendered too small to detect by the unfolding process or could reflect the early loss of protein contacts necessary for the vibronic states causing the further peaks.

The idea of early spectral changes is backed up by later experiments recording the spectra of pSRII in low concentrations of SDS, over timescales with no noticeable unfolding. As can be seen in Figure 4.8a, the position of the pSRII main absorbance peak shifts from around 496nm in the native state and up to 0.1% SDS, to 505nm at 0.3% SDS. This change can be seen both in the coefficients for the peak fitting (as discussed in section 4.3.4), and by taking the local maximum around 500nm. A decline in the areas of fitted shoulder peaks can also be seen in figure 4.8b. The change seems to plateau from 0.2% to 0.3% SDS.

These changes make the fitting of unfolding time courses in 0.5% SDS difficult. This is the lowest SDS concentration with significant unfolding, and though some shoulder peaks are visible early on, the decline of these shoulder peaks is not uniform. As can be seen in the fluorescence time courses (see section 4.4.1.4), the rates of unfolding reactions are such that little of the protonated Schiff base intermediate can be observed, so if the same fitting as for

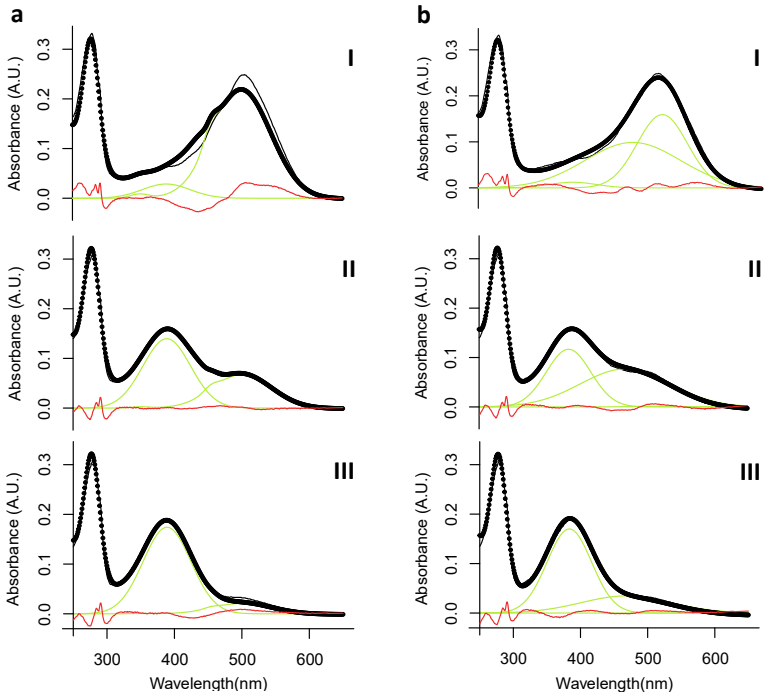


Figure 4.9: Difficulties fitting pSRII unfolding in 0.5% SDS (a) Fitting of pSRII unfolded in 0.5% SDS as one reaction at **(I)** 0 minutes, **(II)** 60 minutes, and **(III)** 120 minutes. **(b)** Fitting of the spectra in (a) to two reactions.

higher [SDS] is attempted, an unrealistically broad 440 nm peak is observed (Figure 4.9b). The k_2 for fitting, which is similar across [SDS] (section 4.4.1.7.2) is also lower for 0.5%.

Improvements can be made by modelling the reaction as a single first order reaction, rather than consecutive reactions as with higher [SDS], as can be seen in the comparison of Figure 4.9a with 4.9b, though with a poor residual at early timepoints. That the greatest magnitude of residual is around 440 nm suggests that the 440 nm peak should be included, it is difficult to reconcile with the overall reaction. As can be seen in figure 4.10, the different models for unfolding in 0.5% SDS have only minor effects on k_1, while the single reaction model brings the parameters for the free retinal peak in line with higher SDS concentrations. As the free retinal peak should not be much affected by SDS, this favours the single reaction model. The centre of the 500 nm peak moves closer to the centre for 0.3% SDS. Due to the improvements made to the overall fitting, and the information from fluorescence time courses, the fitting of a single reaction was chosen.

4.4.1.2: Hydrolysis of the Schiff base

Absorbance of free retinal has a λ_{max} of 387 nm and a deprotonated retinal Schiff base has a λ_{max} of 380 nm. The centre of the peak of the third species modelled in absorbance kinetics varies between 385 nm and 402 nm. If the unfolding of pSRII in 0.5% SDS is modelled as a single reaction, the spread is between 388 and 402nm, as shown in Figure 4.10i. This

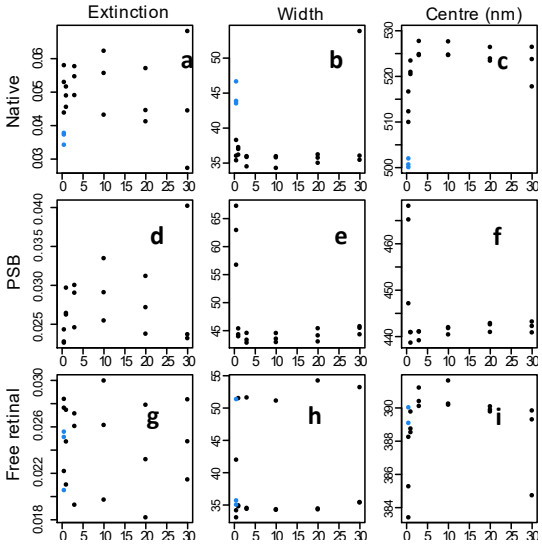


Figure 4.10: Changes in parameters of the combined kinetic model at different concentrations. Top row: 500nm peak parameters: extinction coefficient, width and centre (left to right). Second row: 440nm peak parameters. Bottom row: 380nm peak parameters. Blue points show the parameters of the two peak model of pSRII unfolding.

characteristic confirms the third species in unfolding spectra to be free retinal rather than deprotonated Schiff base.

4.4.1.3: Absorbance kinetics

Rates of transition to 440nm increased with [SDS] nearly from ~0.15 min^{-1} to ~1.05 min^{-1} as SDS increased from 1 to 3%. The 440 to 387 transition did not change with [SDS], remaining around 0.03 min^{-1} (see 4.1.7.1, below).

4.4.1.4: Fluorescence kinetics

Below 1% SDS, only one process was supported by fluorescence time courses. The observed rate of this single process was below 0.03 min^{-1}. Once two processes could be distinguished, k_2 stayed constant at 0.03 min^{-1}, supporting the idea that only once k_1 is higher than k_2 can k_1 be seen, as outlined in section 4.3.3.

4.4.1.5: Fluorescence amplitudes

As well as the kinetic parameters, fluorescence change amplitudes were acquired from the fitting of fluorescence time courses (Figure 4.11d). In the kinetic model, these should reflect the fluorescence changes of the different states. The total fluorescence difference between native pSRII and pSRII without retinal was assumed to be constant relative to the concentration of the sample as the sensitivity of fluorescence quenching to distance is such that retinal released to the bulk solution should have no quenching effect on pSRII tryptophans. For this reason, the decline in A_2 was thought to be accounted for by unfolding occurring in the dead time of the experiment. The ratio of A_1 to A_2 was then used as an indicator of the extent of unfolding by SDS at different concentrations because the extent of loss of quenching shows the separation of retinal from tryptophans before hydrolysis. Initial analysis by linear regression of SDS percentages of 1, 3, 10 and 20% showed a strong positive correlation. From previous work on rhodopsin, the geometry of SDS micelles was shown to have an importance beyond that of just the concentration (Dutta et al., 2010a). To test this, further experiments at 7 and 15% SDS were run (figure 4.11d). The R^2 value of the linear regression decreased from 0.94 to 0.70.

4.4.1.6: CD fitting

Changes in MRE at 222nm over time were used to measure the rates of loss of secondary structure in unfolding pSRII. The same analysis as for fluorescence time courses was carried out on the MRE data. Model selection by AIC consistently chose a single reaction for the data. As the second reaction in fluorescence and absorbance data measures the hydrolysis of retinal, there is no reason to assume a second process. It should be noted that the noise in the data is high because the changes in secondary structure are minor, and model selection only selects for the best model supported by the data. For this reason, the selection of a single process in this case, with high noise, is less meaningful than between fluorescence measurements (see section 4.3.3.1 for details).

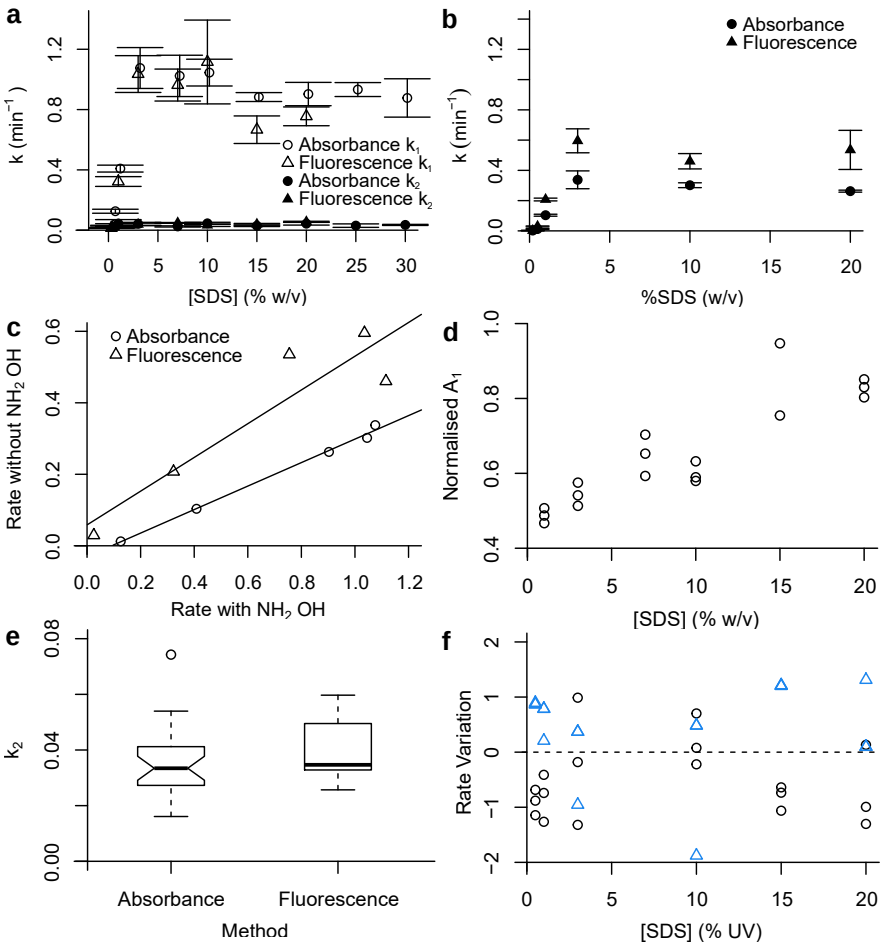


Figure 4.11: Rates of unfolding reactions for pSRII in SDS. (a) Rates for the reactions outlined in the text as measured by absorbance spectroscopy (hollow shapes), and fluorescence (filled shapes). Where there is significant overlap, symbols have been slightly offset in the x-axis. Where two reactions are present, k_1 is shown as a circle, and k_2 as a triangle. Rates are shown as the mean +/- standard deviation for 3 reactions. **(b)** As (a) with hydroxylamine added. Only single reactions were fitted. **(c)** Comparison of rates with and without hydroxylamine. Linear regression was performed between rates for the measurement methods. **(d)** Dependence of A_1 as a fraction of total amplitude for fitting of two curves to fluorescence time courses. **(e)** Comparison of boxplots of k_2 as estimated by fitting of absorbance and fluorescence time courses. No significant difference was found by t-test. **(f)** Comparison of k_1 as compared to the overall average of all estimates at that SDS concentration. Differences shown as standard deviations from the mean at 0. Absorbance values are shown as black circles, fluorescence values as blue triangles.

The rate constants extracted from the fitting to CD data increase with SDS concentration. The rates are lower than the rates measured for tertiary structure changes.

4.4.1.7: Comparison of Absorbance and Fluorescence

Both absorbance and fluorescence data were used to explore the kinetics of pSRII unfolding. Though each experimental system could be sensitive to the same changes, and both output rate constants for two consecutive reactions, these may not be the same across measurements.

4.4.1.7.1: k_2

The dependence of the rate of the hydrolysis of the Schiff base on [SDS] was estimated by ANOVA, and no significant dependence was found for either method of measurement (figure 4.11e). A Welch's t-test between the two was carried out and no significant difference was found. Treating the process as constant across [SDS] and measurement method, the mean k_2 was found to be 0.0391 min^{-1}.

4.4.1.7.2: k_1

No explicit model for the dependence of k_1 upon [SDS] was formulated. It increases sharply towards 3% SDS, as shown in figure 4.11a. If the rates had been maintained, an exponential growth decay model would suffice, but they appear to decline slightly. Instead of fitting to a model, k_1 between measurements were compared by taking the mean of all k_1 at a given SDS concentration and calculating how many standard deviations from the mean each measurement was and observing any differences between measurement methods, as shown in figure 4.11f. Across the range of SDS concentrations, fluorescence tends to be lower than the mean, and absorbance higher. This may be because the techniques are responding to slightly different changes in pSRII, as discussed below.

4.4.1.8: Addition of Hydroxylamine

Hydroxylamine was introduced because in rhodopsin studies it has been used to instantly hydrolyse the Schiff base, and the same property here was intended to simplify the kinetics of the pSRII unfolding reaction by effectively removing the protonated Schiff base from the reaction. Hydroxylamine will react with the Schiff base when it is exposed to the solvent, so, as the helical bundle is loosened, instead of a protonated Schiff base absorbing at 440 nm forming, a retinal oxime absorbing at 380 nm should form. In the absorbance spectra this was achieved; the 440 nm peak present in hydroxylamine free spectra was absent. However, the shoulder peaks discussed in the refolding section of the methods development in this chapter, absent in the spectra of high SDS denatured pSRII, appear in these unfolding spectra. The same procedure used to determine whether these peaks are in a constant ratio with the

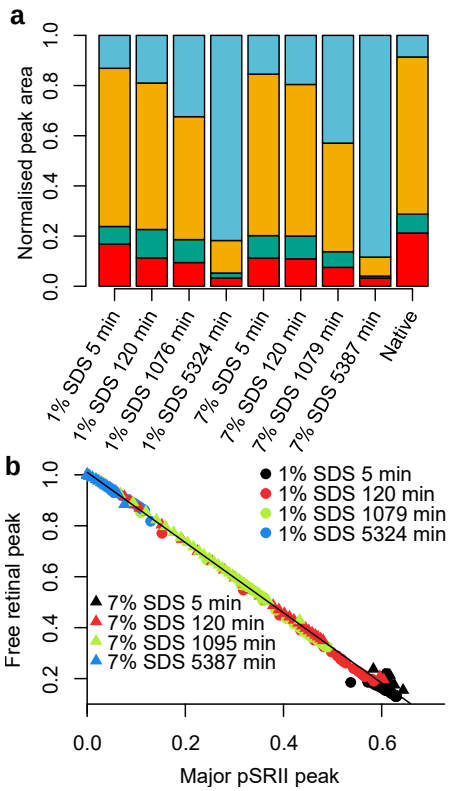


Figure 4.12: Refolding of pSRII by dilution of SDS. **(a)** The maximum recovery of pSRII unfolded in SDS for different times, as assessed by the minimum area share taken by the free retinal peak. Spectra were fitted into four retinal peaks: 372nm (aqua), 496nm (yellow), 461nm (green), and 435nm (red). **(b)** Scatter plot of the area share of the 496nm and 372nm peaks for all spectra. The black line shows the linear regression for the shown values, and has an R^2 value of 0.9973 on 286 degrees of freedom.

main peaks used in refolding spectra was used here, and the increase in the 380 nm peak was determined to be a sufficient measure of pSRII unfolding.

In fluorescence spectra, model selection by both AIC and cross-validation chose different time courses at the same [SDS] as single and double reactions. Both methods made the same model selection for each time course. It was decided that all reactions should be modelled as single processes due to the information supplied by absorbance spectra (see Discussion for details).

The rates measured upon addition of hydroxylamine were slower than those in SDS only unfolding (figure 4.11b). The relative rate change was different between absorbance and fluorescence (Figure 4.11c). The correlation between mean rates at each concentration of SDS +/-Hydroxylamine was estimated by linear regression. As can be seen in figure 4.11c, the rates as measured by fluorescence are approximately halved, and those measured by absorbance are divided by three.

4.4.1.9: Refolding

The fraction of pSRII that could not be recovered by dilution of SDS was more dependent upon the time spent unfolded than on the percentage of SDS, as can be seen in Figure 4.13a. There was a small effect of SDS concentration on the recovery efficiency, though without replication and adding further SDS concentrations, it is uncertain whether this is simply due to the extra time spent in the unfolded state from faster unfolding. The increase in unrecoverable pSRII could be modelled for each concentration as a first-order reaction, with half-lives of 1d 16h 23min for 1% SDS and 1d 6h 10 min for 7% SDS.

The rates of refolding as measured by decrease in free retinal peak could be fit to single exponential decay curves. However, the rates for 7% SDS could be better fit, as assessed by AIC, as the second reaction of two consecutive first-order reactions, as shown in Figure

4.13b. The nature of the first of the two reactions cannot be ascertained from this data; possible candidates are slower mixing of micelles with higher SDS, or compromised mixing of retinal due to higher viscosity. As these differences were difficult to explain, and inconsistent between SDS concentrations, time courses were modelled as a single reaction.

Fitting of the refolding of pSRII showed that dilution of SDS is sufficient to refold, as the vibronic shoulder peaks of the chromophore are restored instantly with the main peak. It also allowed the construction of chevron plots (by Yi Lei Tan, University of Cambridge) for analysis of the combined kinetics of folding and unfolding (Curnow and Booth, 2007).

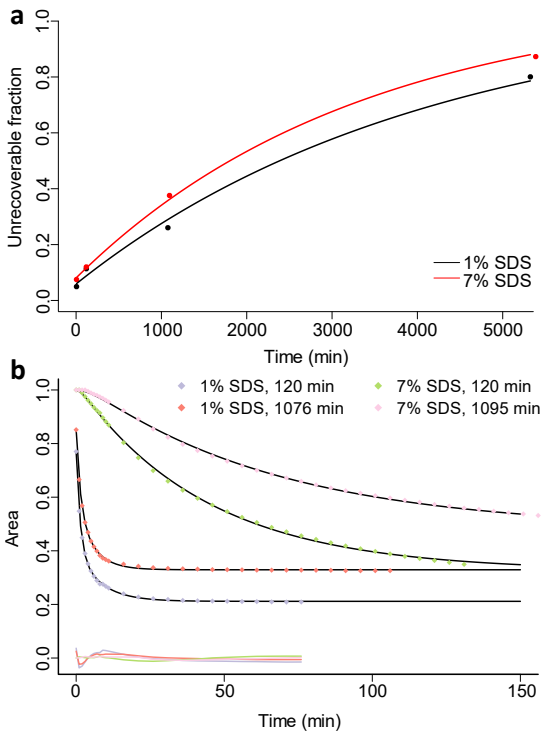


Figure 4.13: Kinetics of pSRII refolding. (a) Scatter plot of the unrecoverable fraction of pSRII unfolded in 7% (red) and 1% (black) SDS. These were fitted to single exponential growth decay curves with rate constants of 3.0 and 2.4 x 10^{-4} min^{-1} respectively. **(b)** Scatter plots of the decrease in free retinal for pSRII refolding. Curves from 1% SDS timecourses are fit to a single exponential decay curve, and those for 7% SDS to the kinetic equation for the second of two consecutive first-order reactions. Residuals shown bottom left.

4.4.2: Analysis of pSRII NMR spectra

4.4.2.1: Principal component analysis

2D HSQC spectra were recorded of pSRII unfolded in different concentrations of SDS up to 30% in order that the environments of different regions of pSRII at different degrees of unfolding might be compared. There were two problems with this approach found after data collection. The first is the volume of data; each of the 239 residues is labelled. Though not all appear as peaks in the dataset, analysis of each residue is inefficient. As concentration of SDS increases, micelles increase in size (Otzen, 2002), and the resultant slower tumbling of pSRII in these micelles increases transverse relaxation. This means a second problem is that the number of peaks decreases with concentration of SDS. Principal components analysis was applied to the data to address both these problems.

Principal components analysis (PCA) is used to reduce the dimensionality of a multivariate dataset while preserving as much of the original information as possible. This can be used to decipher underlying patterns in large datasets, and to cluster samples. It has previously been used with NMR spectra (Konuma Tsuyoshi et al., 2012; Sakurai and Goto, 2007). Another

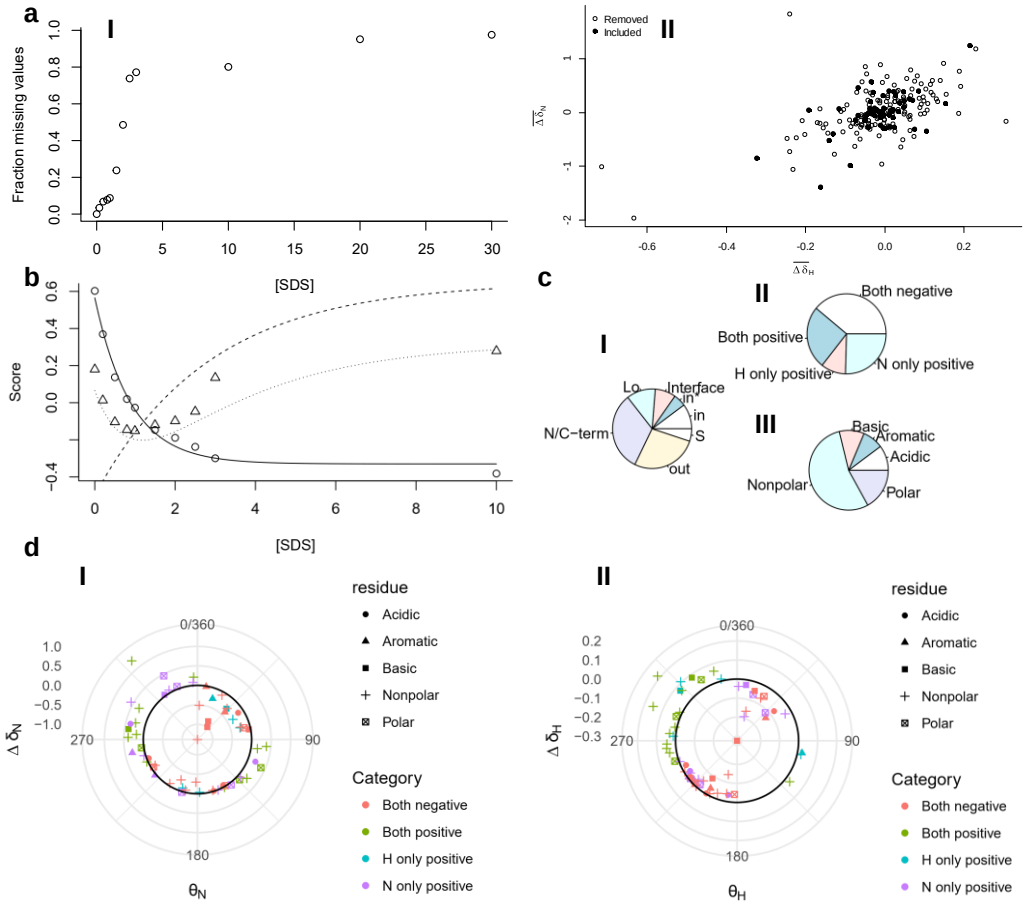


Figure 4.14: PCA of pSRII NMR spectra. (a) (I) Sparseness of data points with increasing SDS. The fraction of total values missing from the matrix of assigned peaks for spectra with different SDS. The maximum is 402/412 at 30% SDS. **(II)** Plot of the maximum chemical shift change of each residue showing the residues in the trimmed dataset are well distributed. **(b)** The scores for the first two principal components for each spectrum. PC1 (R^2 = 0.7485) in circles, PC2 (R^2 = 0.1810, cumulative R^2 = 0.9295) in triangles. The solid line shows an exponential decay curve fitted to PC1 ($y = 0.895e^{-1.12x}-0.33$). The dotted line shows the sum of the exponential decay curve for PC1 and an exponential decay growth curve ($y = 1.15(1-e^{-0.35x})-0.5$) (dashed line) that was fitted to PC2. **(c)** Pie charts of the categories of the 56 residues in the trimmed dataset as divided by **(I)** Domain. The categories are: in the terminal loops (N/C-term), in loops (Lo), at the interface between loops and helix (Interface). Residues in helices are divided into those pointing into the bundle (in), those pointing towards an internal water (in*), and those pointing out (out) (WHAT IS S?); **(II)** Direction of maximum chemical shift change; and **(III)** Chemical characteristics of the residue. **(d)** Polar plots as described in the text for **(I)** Nitrogen chemical shifts, and **(II)** Hydrogen chemical shifts.

useful feature of many PCA algorithms is that many incorporate procedures for adding back in missing data, known as imputation of values (Dray and Josse, 2015).

Each peak from HSQC spectra of unfolded states of pSRII was assigned to a residue (Gautier et al., 2010), and the chemical shift (both ^{1}H and ^{15}N) and the intensity recorded. PCA is, however, sensitive to the scaling of variables, so the representation of different types of measurement must be carefully considered. There are reliable methods of scaling ^{15}N chemical shifts to ^{1}H chemical shifts (Sakurai and Goto, 2007), but scaling intensity is more difficult because the data is of a different kind to the chemical shifts. For the PCA, the spectra were represented as rows in a matrix in which each column held a chemical shift, either ^{1}H or ^{15}N.

4.4.2.1.1: Imputation of values

PCA using an imputation method is useful for the pSRII dataset, where the disappearance of peaks at high [SDS] causes gaps in the matrix. This was the original purpose of PCA suggested for the analysis of pSRII NMR spectra, because if the whole dataset is included, only 40% of the matrix is occupied, and some variables are only represented in one sample. The gaps are also not randomly distributed; their representation increases with increasing [SDS] (figure 4.14a (I)). The power of the PCA increases with the number of variables used, but simply optimising for maximum peaks would involve omitting most [SDS]. Based on the biophysical kinetics data, the decision was made to use the spectra up to 10% SDS, as most of the rate increase has been achieved at that point, and most residues are missing from higher concentrations of SDS.

For this reduced dataset, variables were ranked according to how many spectra they had values for, then variables were removed starting with those present only in the native spectrum and continuing until NA values (empty spaces) represented less than 10% of the matrix, leaving 56 residues of the original 118 in the data set. In practice, this left 6.61% missing values in the dataset due to the block removal process. To make sure that a fair representation of the chemical shift changes remained the maximum chemical shift changes of the whole and the reduced datasets were compared (see figure 4.14a(II)). There appears to be no particular pattern of omission.

4.4.2.1.2: Principal components as a function of [SDS]

Probabilistic PCA (ppca) estimates the principal component axes using an expectation maximization algorithm, and imputes values well compared with other PCA methods. As the gaps in the matrix were one of the key concerns, this method was used to perform PCA.

There are various methods for choosing the number of principal components to return from the analysis (Cangelosi and Goriely, 2007). Many of these are complex to implement and carry with them theoretical commitments to one or another statistical outlook. An easy rule of thumb is to simply return as many components as cumulatively account for a percentage of the total variance. A threshold which often returns component numbers close to the more complex methods is 80%. In theory, one fewer principal component as you have variables can be returned, but this is not useful for most applications. The first two principal components, the vectors from PCA which show the correlation of the input variables, of the reduced dataset preserved 92.95% of the variance, so no further principal components were calculated. As the samples lie on a continuous variable external to the PCA, rather than plotting the principal components against one another, as is often used for categorical data,

the principal components were plotted against [SDS]. The first principal component is fitted well as an exponential decay against [SDS], and the sum of the principal components as an exponential growth decay, so the second principal component is fitted as the sum of these, as shown in figure 4.14b. The functions to which the principal components were fitted can be found in the figure legend.

4.4.2.1.3: Loading analysis

It is very difficult to link the output of PCA to physical phenomena. Some variables contribute strongly to one or another principal component, and which principal component correlates with which variables can point to a unifying explanation. This is quantified as the loadings of each variable in the PCA. To visualise how these are distributed in comparison with transitions in the 2D spectra, each residue in the reduced dataset was plotted in a polar plot.

In a polar plot, data points are represented not with numerical coordinates as in the Cartesian system, but as vectors with a direction relative to the origin and a magnitude. Each can be converted to the other using simple trigonometry. To combine the PCA information from the first two principal components and the chemical shift data, the principal components were converted from Cartesian to polar coordinates. The angle of the polar coordinate is then taken, but the magnitude of the vector is changed to be the maximum chemical shift change for that peak. A circle at 0 was drawn around the origin to make the visualisation easier, as the chemical shift change of peaks could be positive or negative. This endeavoured to show whether how much a residue correlated with one principal component (PC) relative to the other was dependent. As the values used for the PCA are separated into the hydrogen and nitrogen chemical shifts, polar plots of this type were constructed for both (figure 4.14d). For each, there appears to be some correspondence between the direction of chemical change and the direction of correlation with the PCs. Particularly evident in the polar plot for nitrogen values, if a residue correlates positively with one PC, and negatively with another (making its θ between 90 and 180 or 270 and 360), its chemical shift change is more likely to be positive, and the opposite is true if it correlates either positively or negatively with both PCs. The axis of this effect is not quite diagonal, nor is it symmetrical, with more variables correlating negatively with PC1 and positively with PC2 than the inverse. This is more marked with hydrogen values. Residues are more likely to be positive or negative in chemical shift change for both H and N than for one and not the other (figure 4.14c (II)), so it is difficult to tell whether this has any bearing on the distribution. This approach alone did not link the principal components to an underlying physical phenomenon.

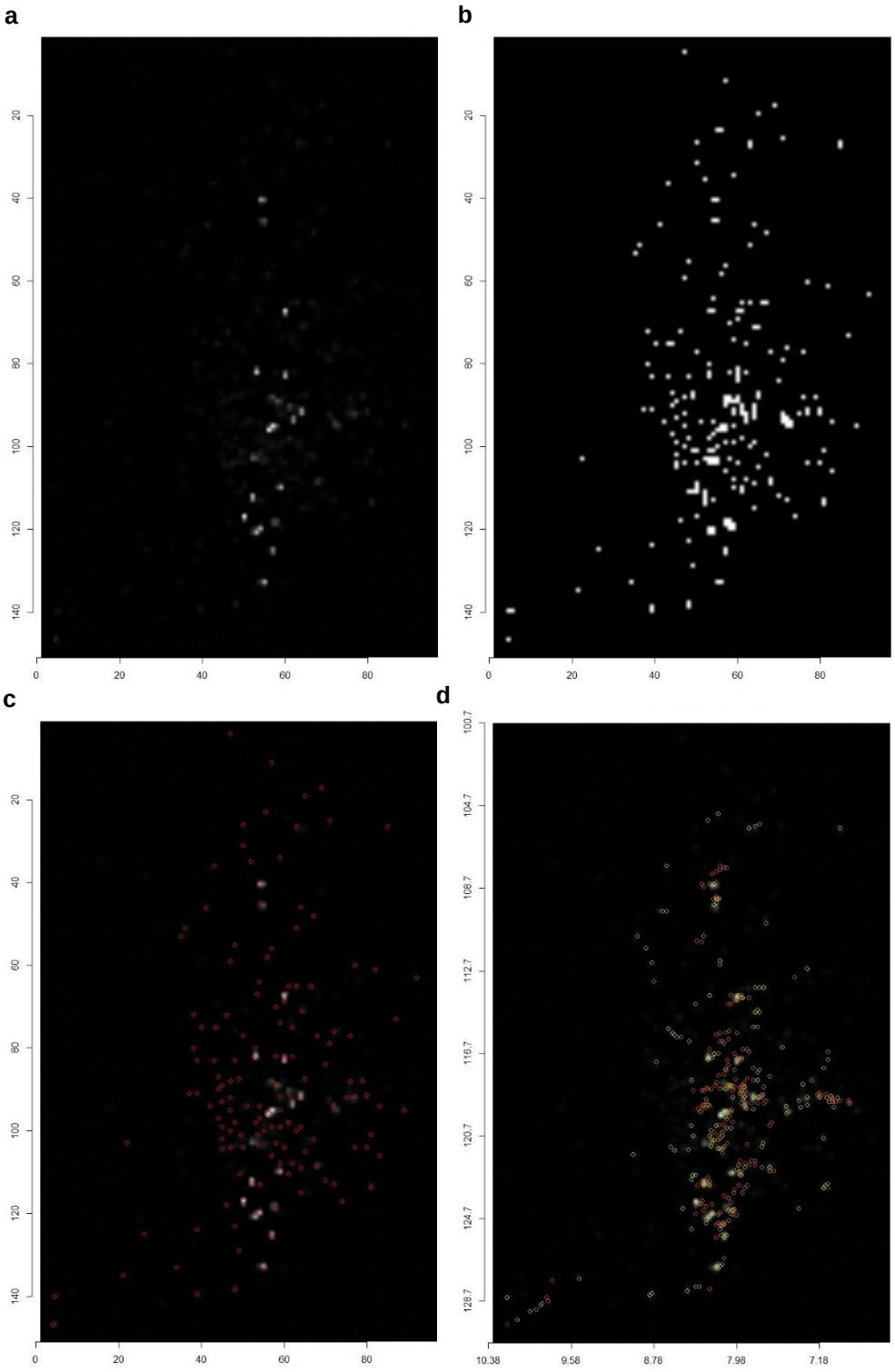


Figure 4.15: Image analysis of pSRII spectra. (a) Bucketed NMR spectrum as a pixel image, with intensity as greyscale values. **(b)** All pixels in the top 1% of values highlighted. **(c)** The centres of each blob of pixels highlighted as red circles. **(d)** Blob centres from 0% to 30% SDS spectra (on continuum from green to red) superimposed on the native spectrum.

4.4.2.2: Other strategies for chemical shift change analysis

4.4.2.2.1: Image analysis

One of the hopes for PCA of NMR of unfolded pSRII was that it would be able to compensate for the disappearance of peaks at higher concentrations of SDS. This aim was only partly

achieved because of the limits of imputation. Another possible avenue was to treat the spectra as images (figure 4.15a) and apply image analysis techniques. Preliminary attempts using "BLOB analysis" (see section 2.3.4 for details) met with some success. As can be seen in figure 4.15, the image analysis technique was able to identify peaks in spectra. An example of peaks identified across spectra is shown in figure 4.15d. The low resolution of the bucketed spectra provided prevented the useful comparison of peak coordinates between blob analysis and the assigned spectra.

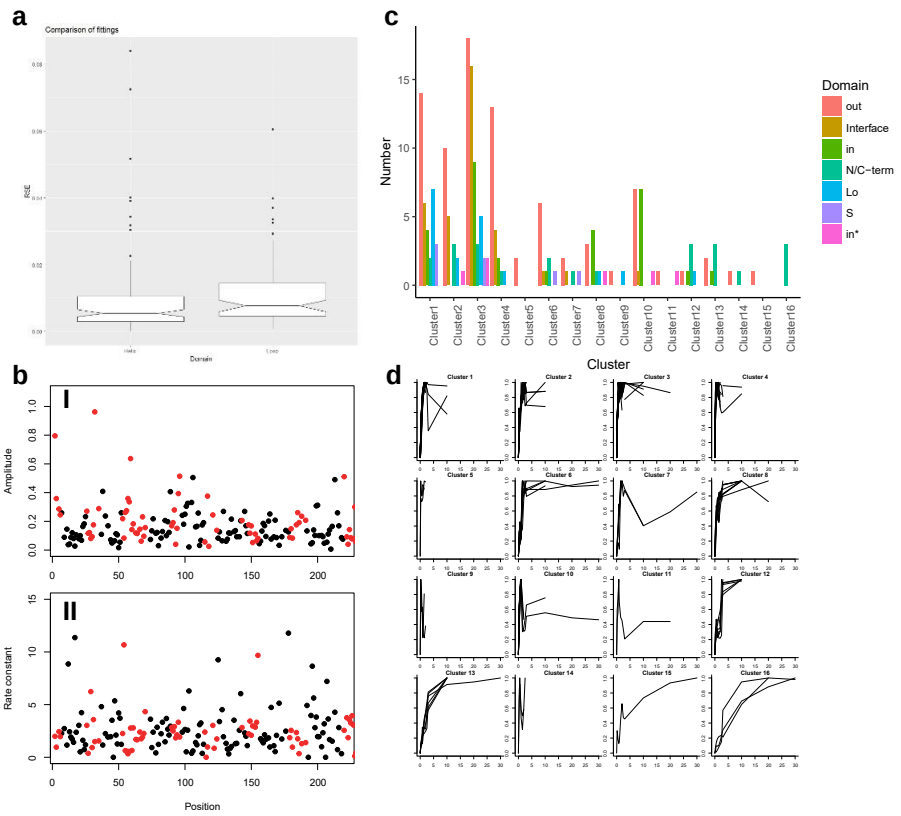


Figure 4.16: Clustering of residues by chemical shift change (a) RSE of fitting a single exponential growth decay curve to the chemical shift change profiles of residues, divided into helix and loop residues. **(b)** Scatter plots of the fit parameters (amplitude **(I)** and rate constant **(II)** for all residues successfully fit to a single exponential curve. Helix residues in black, loop residues in red. **(c)** All residues clustered by normalised chemical shift change as a function of SDS. The hierarchical clustering was split into 16 clusters. A χ^2 test of the clusters against domain has a p-value of 3.09×10^{-3}. **(I)** A barplot of the number of residues of a given domain (by colour) in a given cluster. **(II)** The normalised profiles of the residues used for clustering divided by cluster.

4.4.2.2.2: Fitting

The magnitude of chemical shift changes generally increased with increased SDS. In most cases, the pattern of changes could be fitted with an exponential growth decay. A small number (12 in helices and 6 in loops) failed to fit entirely. The root square errors of the fits were used as a measure of goodness of fit, and outliers more than two standard deviations from the mean of RSE in the upper bound were examined further. For helices, this meant fits with an RSE > 0.0212, and for loops 0.0272 (figure 4.16a). For the good fits, there was no significant difference between helices and loops for the amplitude of increase against SDS

concentration. However, when each parameter was plotted against the residue's position in primary structure, there appeared to be some periodic variation in amplitude (figure 4.16b), but no satisfactory formal profile could be established.

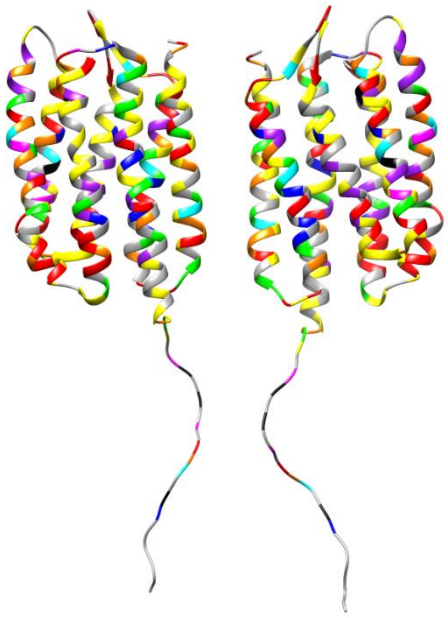


Figure 4.17: pSRII structure coloured by cluster. Clusters with fewer than five members are not coloured. Cluster 1 = red, Cluster 2 = orange, Cluster 3 = yellow, Cluster 4 = green, Cluster 6 = cyan, Cluster 8 = blue, Cluster 10 = purple, Cluster 12 = magenta, Cluster 13 = black.

The shapes of the poorly fitted residues were compared by first normalising them to values between 0 and 1 by dividing the magnitude of change at a given [SDS] by the maximum magnitude for that residue. Each residue was then treated as a point in n-dimensional space, where n is the number of SDS concentrations at which the residue was observed, and its magnitude in each dimension is the magnitude of chemical shift change at that concentration. The values were clustered by the daisy algorithm (Introduction to "Finding groups in data," 2008) using Gower distances (Gower, 1971). They were then separated into 4 groups (4 was chosen arbitrarily) in hope that better models for the patterns of their change could be defined, but without success.

Once this had been done, the clustering process was repeated for all residues with a peak in at least two spectra, with 8 clusters. A χ^2 test was performed on the clusters to see if the classification could distinguish residue chemical properties or domains. No significant classification in chemical properties was achieved, but the χ^2 value for domain classification was 82.44 on 42 degrees of freedom, with a p-value of 1.93×10^{-4}. The procedure was repeated with 16 clusters, which had a χ^2 value of 131.07 on 90 degrees of freedom, with a p-value of 3.09×10^{-3}. The populations of these clusters are shown in figure 4.16c (I). The profiles of the residues used for clustering are shown in figure 4.16c (II), divided by cluster. The larger clusters have varying profiles, and some clusters have very few members. However, clusters 10 and 12 have distinct profiles and are well populated (16 and 6 respectively).

4.5: Discussion

Kinetics of unfolding

To compare the folding of pSRII with MR and bR, an experimental unfolding system had to be devised. Initial experiments with SDS unfolding, the most successful reagent for MR unfolding, suggested that pSRII would also be informatively unfolded. A screen of other reagents and conditions was carried out by collaborators at the University of Cambridge (paper in preparation), but the use of SDS achieved the greatest unfolding without aggregation. Though a similar experimental setup was used for pSRII and MR unfolding (Dutta et al., 2010a), absorbance time courses of pSRII unfolding showed very different kinetics, with much slower disruption of retinal-protein contacts, and faster hydrolysis of the Schiff base (figure 4.2), making analysis of the reactions on the same time scale possible. Devising a combined kinetics and decomposition model (outlined in section 4.3) allowed the separation of two unfolding processes: loss of specific protein contacts and hydrolysis of the Schiff base. The early changes in absorbance spectra are indicative of events outside the time resolution of the current experiments which may show further differences between unfolding of pSRII with different SDS concentrations. BR is closely related to pSRII, with an absorbance maximum at 570 nm. Specific residues are responsible for the blue shift of pSRII compared with BR, including asparagine 72. The fact that in early spectra the λ_{max} of the native-like states of pSRII in low [SDS] are shifted to the red and the shoulder peaks disappear suggests that the interaction of this residue is disrupted earlier on compared to other retinal binding pocket residues, but the fast kinetics are beyond the reach of the experimental setup.

The separation of processes provided by decomposition of absorbance spectra also allowed comparison with fluorescence time courses. After finding a single exponential curve was inadequate for modelling the increase in fluorescence upon unfolding, the knowledge of kinetics taken from absorbance informed the modelling of this as a superposition of two kinetics equations, which was then vindicated by the fact that the second reaction of both methods had the same rate, constant across SDS concentrations (figure 4.11e). That the first reaction was similar, but not exactly the same between measurement methods (figure 4.11f) does not contradict these models, as the changes absorbance spectroscopy is sensitive to are not the proximity of tryptophan residues, which is detected by fluorescence. The rates recorded by fluorescence are generally lower, but perhaps more interesting is that the dependence of these rates upon SDS concentration is also slightly different. For both, the rate declines beyond 3% SDS. When this was first observed in kinetics determined by

absorbance, it was thought that perhaps mixing was compromised at very high SDS due to viscosity. It is possible that this is the explanation for both absorbance and fluorescence experiments, and that differences in the experimental setups account for the differences in recorded rates. It is also possible that the rate differences are due to artefacts of the fitting process, but then it would be reasonable to assume a similar effect on rates of the second reaction. Had the rates of the second process for each been different, the models would need revision.

Like the revelation of early absorbance spectral changes revealed by decomposition, the modelling of fluorescence time courses revealed more than kinetic information. That the fluorescence change between the as a proportion of the total fluorescence increase is larger at higher SDS shows that loosening of the helical bundle is not only faster but to a greater extent as SDS increases. Whereas the kinetics reach a peak at 3% SDS and slightly decline at higher SDS, the relative share of the first curve continues to increase. Using kinetic models means that this is not just due to increased rates, as might be observed with models of parallel processes. Without this decomposition, such a conclusion could not have been reached.

Unfolding in the presence of hydroxylamine was intended to simplify the fitting process and confirm the rates of the first reaction because in MR, formation of the retinal oxime is instantaneous with the formation of the protonated Schiff base. However, in pSRII this is not true, and though the overall reaction from native pSRII to release of retinal is increased, the reaction is not as rapid as the loosening of the helical bundle. This caused problems with the fitting of fluorescence time courses. Whether a single or double exponential is a better model at a given SDS concentration more ambiguous than for the same reaction without hydroxylamine. This is likely to be because the two reactions occur at similar rates, which has a similar profile to a single reaction.

With these observations, it is somewhat surprising that, in absorbance spectra, no protonated Schiff base is observed. Indeed, the shoulder peaks are observed at SDS concentrations where they are absent in spectra without hydroxylamine. This implies that hydroxylamine reacts with retinal in unfolding pSRII quickly enough to prevent buildup of protonated Schiff base or whatever intermediates have altered spectra (see section 4.4.1.8, Figure 4.11b), but somehow slowly enough that the overall reaction is not as fast as the first reaction in the two consecutive reactions observable without it. It is also interesting that the effects on the rate caused by hydroxylamine might be different as measured by fluorescence

and absorbance, illustrating the differences between the loss of interactions causing the change to protonated Schiff base and the loosening of the bundle which moves the tryptophans away from retinal. This may not be a robust conclusion because of the difficulties decomposing fluorescence time courses.

Fitting of CD time courses was less informative. All time courses were better modelled by a single exponential curve than two. That the secondary structure should deteriorate in a single process would be unsurprising, but it is possible that the noise in the measurements might be covering up a second process, so this result is not conclusive.

Refolding

Decomposition of the absorbance time courses for refolding was also informative. First, it demonstrates that pSRII is already properly refolded without retinal with two observations. First is that no protonated Schiff base forms, as no 440 nm peak arises, and second is that the native profile of shoulder peaks, dependent on the native vibronic state of retinal with pSRII, is instantly established. It also allowed an accurate determination of the proportion of pSRII refolded and shows that the time spent in the unfolded state, rather than the extent of unfolding, is the greatest determinant of whether pSRII will refold. The confirmation that a native profile is instantly established allowed the automation of the refolding kinetics fitting, a script for which is currently being used by collaborators in Cambridge to explore refolding. The automated refolding assumes a single exponential curve models the refolding reaction, which generally serves well. The residuals do have a small, early discrepancy, similar to the pattern observed in the fitting of single processes to the consecutive first order reactions in fluorescence. These are larger in 7% SDS than in 1% SDS. Fitting two exponential decays to the 1% SDS data improves the fit as measured by AIC, but not for 7% SDS. However, using the rate equation for the second of two consecutive reactions fits better for 7% SDS. As no 440 nm peak is observed in this data, it appears that the protein has refolded before binding retinal. The improvement in fitting at 7% by using the rate equation might suggest the refolding is slightly slower at 7% SDS than 1%, where even a single exponential decay makes a better fit, or it could reflect a slower mixing of the c7 into the SDS micelles, which could be the first of two reactions which is invisible to absorbance spectroscopy.

Difficulties in fitting refolding time courses led to the development of an automatic peak picking algorithm. The fact that skewed peaks were present in the spectra prevented the peak picking automation being useful here. However, the algorithm still showed its usefulness to identify small peaks in a noisy background environment. The shoulder peaks in pSRII were not detected well by differential peak picking methods, so it was hoped that by

“oversmoothing” the data and subtracting this from the dataset they could be distinguished. In hindsight, a better optimised method using differentials derived from Savitzky-Golay filtering of the data, or some combination of the two, may be useful, as was found for the fitting of Ce6 spectra. This method struggled with smaller, broader peaks, but the Savitzky-Golay filter is highly customizable, so a future implementation may perform better. If the current strategy remains useful, there are improvements that can be made, especially in the starting estimate stage.

Noise registers as small peaks at this stage. An extra smoothing filter step between oversmoothing and applying the rolling disc is likely to resolve this issue. A second issue is the distortion of larger peaks and occasional loss of a broader peak during the application of the rolling ball. Again, this does not necessarily find representation in the final fitting but puts an undue burden on later stages of the estimation. Using different parameters for the rolling disc can mitigate these issues, but as yet no general optimal solution has been found.

NMR of unfolded states

HSQC NMR spectra provide a wealth of information that may not be amenable to unaided human analysis. The use of PCA on HSQC spectra of pSRII unfolded states was originally suggested as a method for imputation of missing values where peaks disappear from the spectra. Imputation methods can only add so many values when the dataset is sparse, so this aim could not be met. However, the PCA was useful in that the first two principal components, which account for 93% of the variance of the data, follow regular patterns across SDS concentration. It is important to note here that the PCA does not “know” that the different spectra are taken at different SDS concentrations, so this represents the identification of two underlying SDS dependent processes observed in the data. The physical nature of principal components is difficult to identify, but the analysis of loading values using polar plots shows that such an identification may be possible. This might also be explored using the clustering of chemical shift change profiles carried out here, which was sensitive to the domains of residues.

Summary

Building on the work on MR unfolding, an SDS unfolding protocol was established for pSRII. The differences in the kinetics of the stages of unfolding meant that the pSRII system lent itself well to more sophisticated analysis of absorbance and fluorescence data, allowing the development of a kinetic model which enabled the extraction of further information relating to specific changes. It was also reversible, and decomposition of absorbance spectra show that this refolding is not retinal dependent. Detailed analysis of 2D NMR spectra of the

unfolded states of pSRII revealed patterns in the data not available to human observation, which may be developed to show whether pSRII strictly follows the two-stage model for folding, or also requires proper long-range interactions, as suggested by FIRST analysis.

The unfolding and refolding reactions of pSRII, analysed by biophysical methods, form a new model system for the folding of helical membrane proteins. The ways this system differs from the folding of bR and MR will provide a new avenue for determining the factors which contribute to different mechanisms of folding. The methods developed here of separating different contributions to time courses of absorbance spectra and fluorescence emission may also be of interest to those investigating kinetics through such observations. The analysis of NMR spectra with PCA is not new (Konuma Tsuyoshi et al., 2012; Sakurai and Goto, 2007), though previous use of the other methods applied to the NMR spectra have not been observed in the literature.

CHAPTER 5: FURTHER DEVELOPMENT OF RHODOPSIN AS A MODEL SYSTEM FOR NMR STUDIES OF FOLDING AND MISFOLDING

5.1: Summary

Mammalian rhodopsin is a well characterised model system for an important family of proteins: the class A G-protein coupled receptors. Its denaturation by SDS yields unfolded states which do not aggregate, allowing the characterisation of these states (Dutta et al., 2010b, 2010a). This makes it a model system for the folding of helical membrane proteins (MP), an understudied field due to the difficulties of working with membrane proteins. Unlike other helical MP folding model systems, residual structure is found in unfolded states, here centred around the extracellular (EC) domain. Samples of rhodopsin with isotope labelled methionines were prepared for nuclear magnetic resonance (NMR) spectroscopy due to the favourable relaxation characteristics of the methyl group of methionine. Samples were prepared in both dodecyl maltoside (DM), the standard detergent for rhodopsin experiments, and lauryl maltose neopentyl glycol (LMNG), which has properties favourable for NMR. These properties were overshadowed by difficulties with sample preparation. Preliminary assignments of the peaks in HMQC spectra were made, and a rationale for using thiol specific spin labels to aid assignment was explored.

This development of rhodopsin as a model system is applicable to future studies of folding and has made possible the NMR study of rhodopsin misfolding in the P23H mutant (chapter 6).

5.2: Introduction

5.2.1: Rhodopsin as a folding model

As discussed in the introduction to chapter 4, the SDS unfolded states of rhodopsin combine a high degree of secondary structure loss with a lack of aggregation (Dutta et al., 2010b). These states were investigated with a variety of both time-resolved and steady-state techniques (Dutta et al., 2010a). In the original investigation, the changes measured were global. A combination of NMR and electron paramagnetic resonance (EPR) experiments were later conducted on rhodopsin denatured in both different concentrations of SDS and a combination of 3% SDS and 8M Urea (3S8U).

The NMR data showed qualitatively that the cytoplasmic domain of rhodopsin becomes progressively more flexible with the addition of SDS, even from 0.05% SDS. The transmembrane and extracellular domains lose both secondary and tertiary structure, but

retain some rigidity even at 30% SDS, in the maximally unfolded state. It has been argued that proteins can never be in a true random coil (Shortle and Ackerman, 2001), and that, for membrane proteins, secondary structure is likely to remain in disordered states (Kjaergaard, 2015). It has also long been observed that SDS can induce non-native helix formation (Grourke and Gibbs, 1967) These experiments allowed a higher level of resolution to earlier experiments and 1D ^{1}H NMR measurements carried out in the same work, discriminating between domains. The EPR experiments introduced labels to individual residues in the cytoplasmic and extracellular domains and showed the same general pattern. However, the single residue resolution also showed that the residual structure was not retained across the whole extracellular domain and that T108, not included in the predicted folding core (Tastan et al., 2007), lost rigidity at low concentrations of SDS. There are drawbacks to EPR, however. The label introduced is bulky, and introducing it requires mutation.

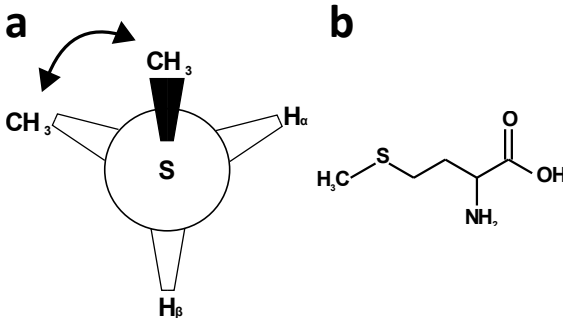


Figure 5.1: Estimation of methionine chemical shifts. (a) View of methionine, showing the angle of trans/gauche rotamerism. **(b)** Structure of methionine.

5.2.2: Rationale for labelling rhodopsin with methionines

Selectively labelling methionine residues reduces the number of peaks in 2D HMQC (see section 1.1.4.1) spectra compared to full labelling, but the assignment of these peaks without mutagenesis is difficult. Previous work has focused on ^{15}N-labeled amino acids, but the signal intensity was found to be highly variable and overall low, with rapid relaxation (J. Klein-Seetharaman et al., 2002; Werner et al., 2007). Incorporating ^{13}C-labeled amino acids has been used for improved signal in large complexes (Beatty et al., 1996; Tugarinov et al., 2003) including GPCRs (Nygaard et al., 2013). We therefore investigated NMR studies of ^{13}C,^{15}N-methionine labelled rhodopsin.

As with ^{15}N-HSQC spectra, there are chemical shift changes associated with the side chain conformations of aliphatic residues. The methyl group of methionine is very flexible and has three proton resonances. The sulphur connecting the methyl carbon has unique effects upon its resonance, and the *trans/gauche* conformation (Figure 5.1a) has been correlated with specific chemical shift changes by comparing resonances from the BioMagRes database with angles obtained from examination of crystal structures from the protein data bank (Butterfoss et al., 2010).

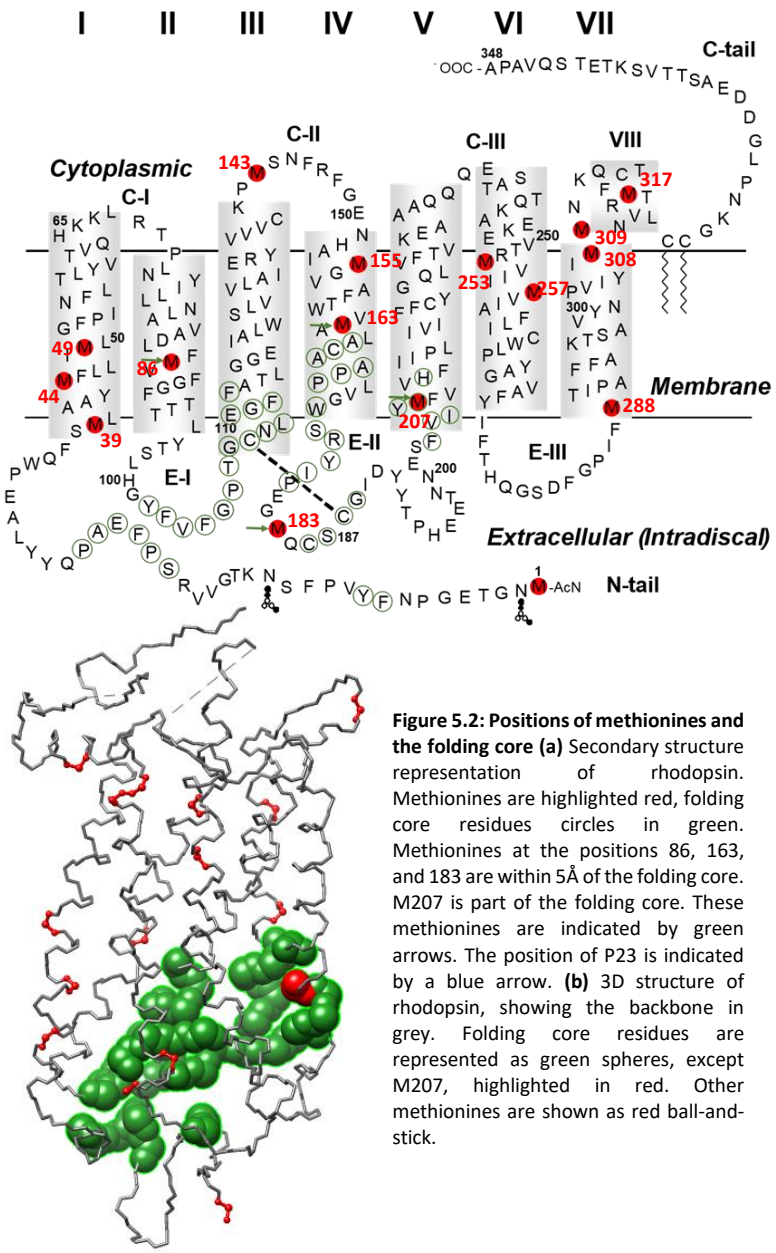


Figure 5.2: Positions of methionines and the folding core (a) Secondary structure representation of rhodopsin. Methionines are highlighted red, folding core residues circles in green. Methionines at the positions 86, 163, and 183 are within 5Å of the folding core. M207 is part of the folding core. These methionines are indicated by green arrows. The position of P23 is indicated by a blue arrow. **(b)** 3D structure of rhodopsin, showing the backbone in grey. Folding core residues are represented as green spheres, except M207, highlighted in red. Other methionines are shown as red ball-and-stick.

Rhodopsin contains sixteen methionines. These are well distributed throughout, with three in loops, five at the interface of helices, and the rest buried in transmembrane helices (figure 5.2a). Importantly for their use in investigating the folding of rhodopsin, four are within 5Å of the folding core. One of these, M207, is within the folding core, as shown in figure 5.2b. The methionines are also well distributed between the faces of rhodopsin. Nine are closer to the extracellular side, meaning five residues are in a similar place to the folding core, without being involved in it, making a useful comparison.

5.2.3: Amphiphiles for solubilisation of Rhodopsin for NMR

n-Dodecyl β-D-Maltoside (DM, Figure 5.3a) solubilises rhodopsin from both rod outer segments and cultured cells and maintains rhodopsin's stability better than many other

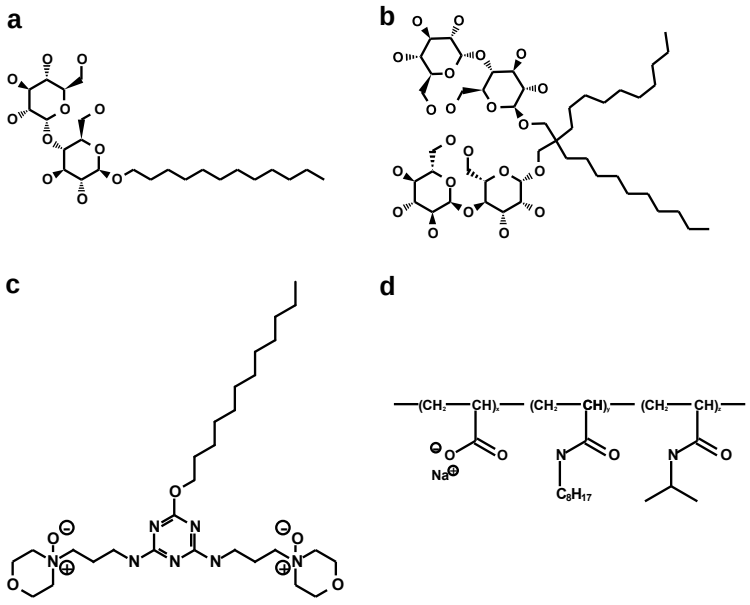


Figure 5.3: Structures of the amphiphiles used to solubilize rhodopsin. (a) n-Dodecyl β-D-Maltoside (DM). **(b)** Lauryl maltose neopentyl glycol (LMNG). **(c)** Triazine based amphiphile courtesy of Andrew Marsh (AM-2). **(d)** Amphipol A8-35 (A835)

detergents (Loewen et al., 2001; Ramon et al., 2003). Solubilisation is necessary for many biochemical and biophysical experiments. It is the detergent used for the other rhodopsin experiments in this thesis and was used to prepare initial NMR samples.

Lauryl maltose neopentyl glycol (LMNG, Figure 5.3b) has been used to solubilise different rhodopsin complexes (Gao et al., 2017; Jastrzebska et al., 2013). Previous NMR work on the β_2-adrenoreceptor showed distinct functional states in LMNG that were not accessible with the protein solubilised in DM (Chung et al., 2012). Better protein stability and lower aggregate mass resulting in better resolution were also reported.

The preceding amphiphiles are traditional detergents. Other compounds with similar properties have been engineered for use in studies of membrane proteins (Bayburt and Sligar, 2010; Popot et al., 2003; Seddon et al., 2004). Amphiphiles based around a triazine core have been designed (Beecham Matthew P. et al., 2013) to incorporate methylamines, which can be kosmotropes, or molecules which stabilize proteins (Kane et al., 2003). It is hoped that the modifiable groups on these amphiphiles can be tailored for particular purposes for particular proteins, for example, incorporation of kosmotropes improves crystallization of membrane proteins, and other modifications have been used to inhibit aggregation (De Bernardez Clark et al., 1999). One of these (molecule 19 in the above

reference, referred to here as AM-2, Figure 5.3c) including morpholino and dodecyl groups was the kind gift of Andrew Marsh (University of Warwick).

Amphipols are unlike traditional detergents. Rather than being small, well defined molecules, amphipols are large polymers with hydrophilic backbones derivatized with hydrophobic groups (Tribet et al., 1996). Amphipols with polyacrylate backbones are highly flexible and do not form intramolecular hydrogen bonds, and those with low weights have been used to solubilise membrane proteins, even from inclusion bodies (Dahmane et al., 2009; Flötenmeyer Matthias et al., 2007) and cell-free expression systems (Etzkorn et al., 2013). Benefits of solubilising in amphipols can be greater stability (Bazzacco et al., 2009; Calabrese et al., 2015) and that they bind almost irreversibly to membrane proteins, so no critical micelle concentration must be maintained in the bulk solution as for detergents. One of these amphipols, A8-35 (Figure 5.3d), named in (Tribet et al., 1996) for its approximate mass (8kDa) and proportion of derivatized groups (~35%), was tested for its ability to solubilise rhodopsin.

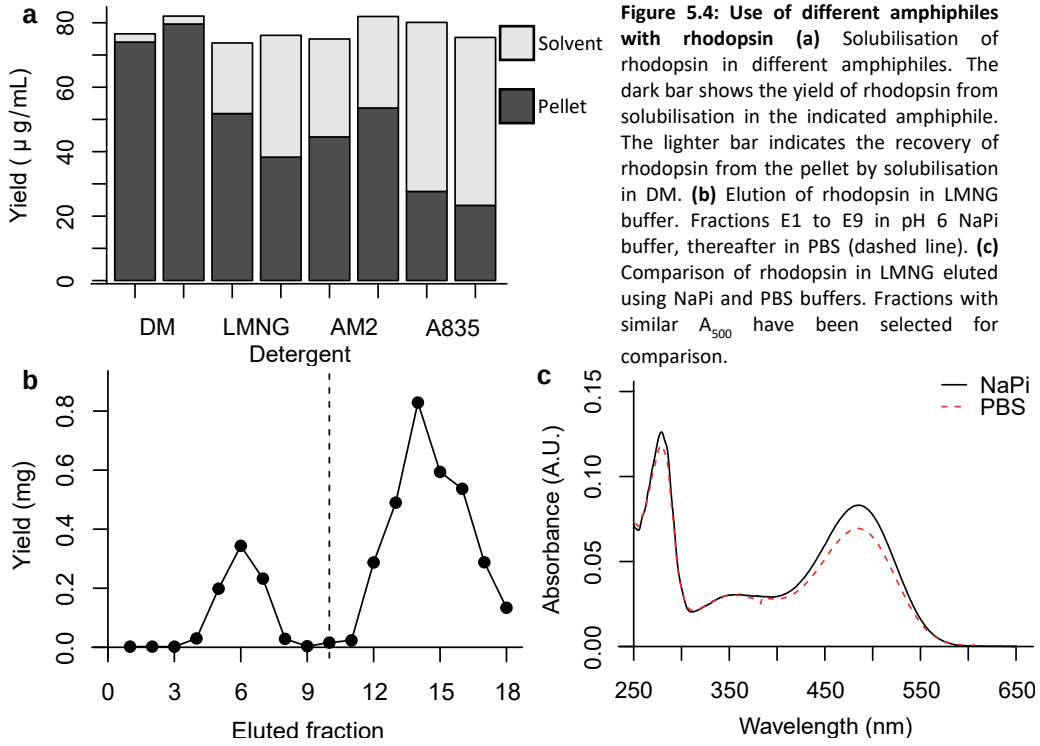


Figure 5.4: Use of different amphiphiles with rhodopsin (a) Solubilisation of rhodopsin in different amphiphiles. The dark bar shows the yield of rhodopsin from solubilisation in the indicated amphiphile. The lighter bar indicates the recovery of rhodopsin from the pellet by solubilisation in DM. **(b)** Elution of rhodopsin in LMNG buffer. Fractions E1 to E9 in pH 6 NaPi buffer, thereafter in PBS (dashed line). **(c)** Comparison of rhodopsin in LMNG eluted using NaPi and PBS buffers. Fractions with similar A_{500} have been selected for comparison.

5.3: Results

5.3.1: Amphiphile screen

The four amphiphiles described in the introduction were tested for their ability to solubilise rhodopsin from rod outer segment (ROS) suspension. The remaining rhodopsin in the pellets from this solubilisation was further treated with DM so comparisons to DM could better be

made. As can be seen in figure 5.4a, DM is best able to solubilise rhodopsin. LMNG and AM-2 are each able to solubilise a little over half the rhodopsin of which DM is capable, and A8-35 around a quarter.

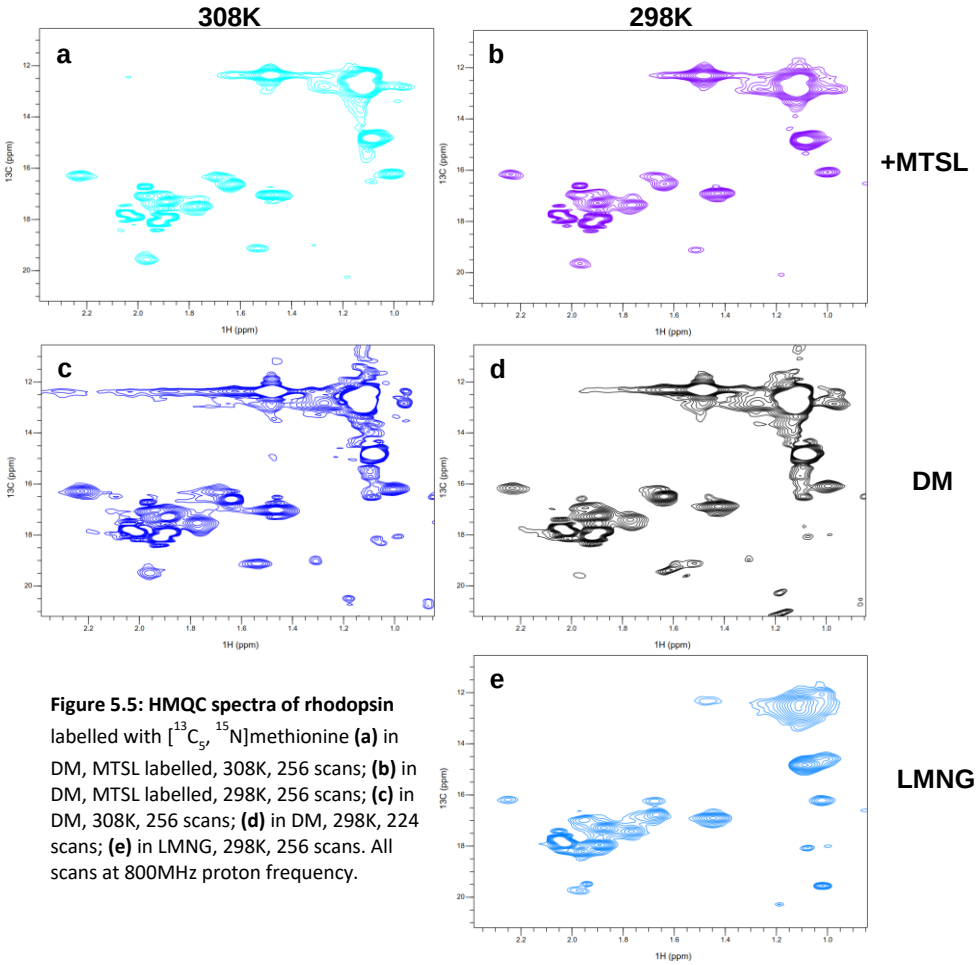


Figure 5.5: HMQC spectra of rhodopsin labelled with [$^{13}C_5$, ^{15}N]methionine **(a)** in DM, MTSL labelled, 308K, 256 scans; **(b)** in DM, MTSL labelled, 298K, 256 scans; **(c)** in DM, 308K, 256 scans; **(d)** in DM, 298K, 224 scans; **(e)** in LMNG, 298K, 256 scans. All scans at 800MHz proton frequency.

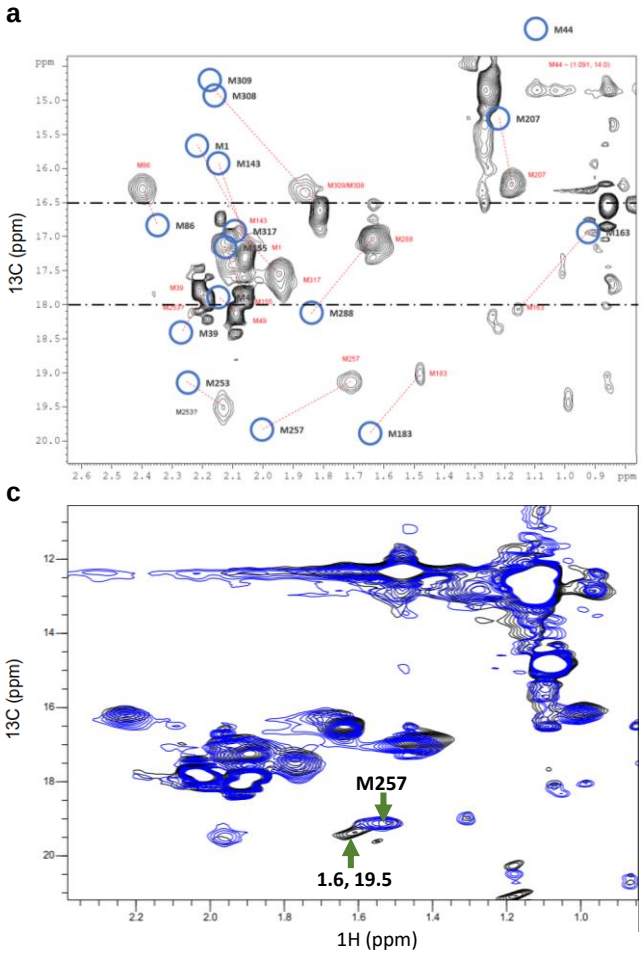
5.3.2: NMR of methionine labelled wild-type rhodopsin in DM

HMQC spectra of rhodopsin in DM were recorded and tentatively assigned using the prediction strategy described in section 5.2.2 (Butterfoss et al., 2010). Predicted values for the peaks are compared with observed values in figure 5.6b. Figure 5.6a shows how these values compare with recorded peaks. Assigning ^{13}C shifts of methionine terminal methyl group (ε) carbons is considered reliable outside the region between 16.5 and 18 ppm. As can be seen in figure 5.6a, many of the recorded peaks fall within this region.

The calculated values are not the only information useful for the assignment of peaks. The relative flexibility of different regions provides information about the expected broadness of peaks.

Spectra were recorded at 298K and at 308K, the latter for the greater resolution at higher temperatures, and the former so that comparison could be made with other samples

potentially unstable at 35°C. The spectra are generally similar, notable differences being the possible loss of a peak at (1.6,19.5) ppm. The peak assigned as M257 is adjacent, so this could merely be the narrowing of that peak.



	Qe shift	Chemical shift	Dihedrals	^{13}C shift
M86	0.215	2.345	1.445	16.8
M39	0.137	2.267	1.677	18.4
M253	0.115	2.245	2.674	19.1
M1	0.084	2.214	-1.288	15.7
M309	0.042	2.172	1.151	14.7
M308	0.027	2.157	-1.185	15.0
M143	0.013	2.143	-1.323	15.9
M49	0.013	2.143	-1.595	17.9
M155	-0.009	2.121	1.490	17.2
M317	-0.041	2.089	1.457	16.9
M257	-0.131	1.999	2.064	19.8
M288	-0.295	1.835	1.630	18.1
M183	-0.489	1.641	-2.214	19.9
M207	-0.913	1.217	1.233	15.3
M44	-1.039	1.091	-1.032	14.0
M163	-1.211	0.919	-1.459	16.9

Figure 5.6: WT rhodopsin in DM with assignments. (a) Overlay of the predicted peak positions for methionines (blue circles) with figure 3(d). Red dashed lines connect the assigned peaks with the theoretical overlay. Dotted-dashed lines indicate the boundaries, 16.5 and 18ppm between which ^{13}C shifts are unreliable. **(b)** Table showing the predictions for each residue. Q_e shift is from a pseudoatom at the centre of He1, He2, He3). ^{1}H shift is from a reference of 2.13 ppm, calculated as in Bundi and Wüthrich, 1979. Dihedrals are and reported as radians, from which ^{13}C shift is calculated from Butterfoss et. al 2010, as outlined in the chapter introduction. **(c)** Overlay of figure 2(d) (black, 298K) with figure 2(c) (blue, 308K)

5.3.3: MTSL labelling of ^{13}C,^{15}N-met rhodopsin

Though many of the peaks in figure 5.6 were assigned from the predictions made from the χ3 rotamer predictions, some of these values are found within the range between 18 and 16.5 ppm and are therefore ambiguous. It was therefore decided to attempt the unambiguous assignment of some peaks using a combination of added paramagnetic relaxation enhancement (PRE) active reagents, which increase the transverse relaxation rate of nearby nuclei (as described in section 1.1.4.1). The effect of these is to diminish the peaks in the space around themselves, so peaks disappearing upon their addition will be close to the reagent.

There are existing protocols which can be adapted to introduce such a reagent to the cytoplasmic face of rhodopsin. S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) is a thiol specific spin label that has previously been

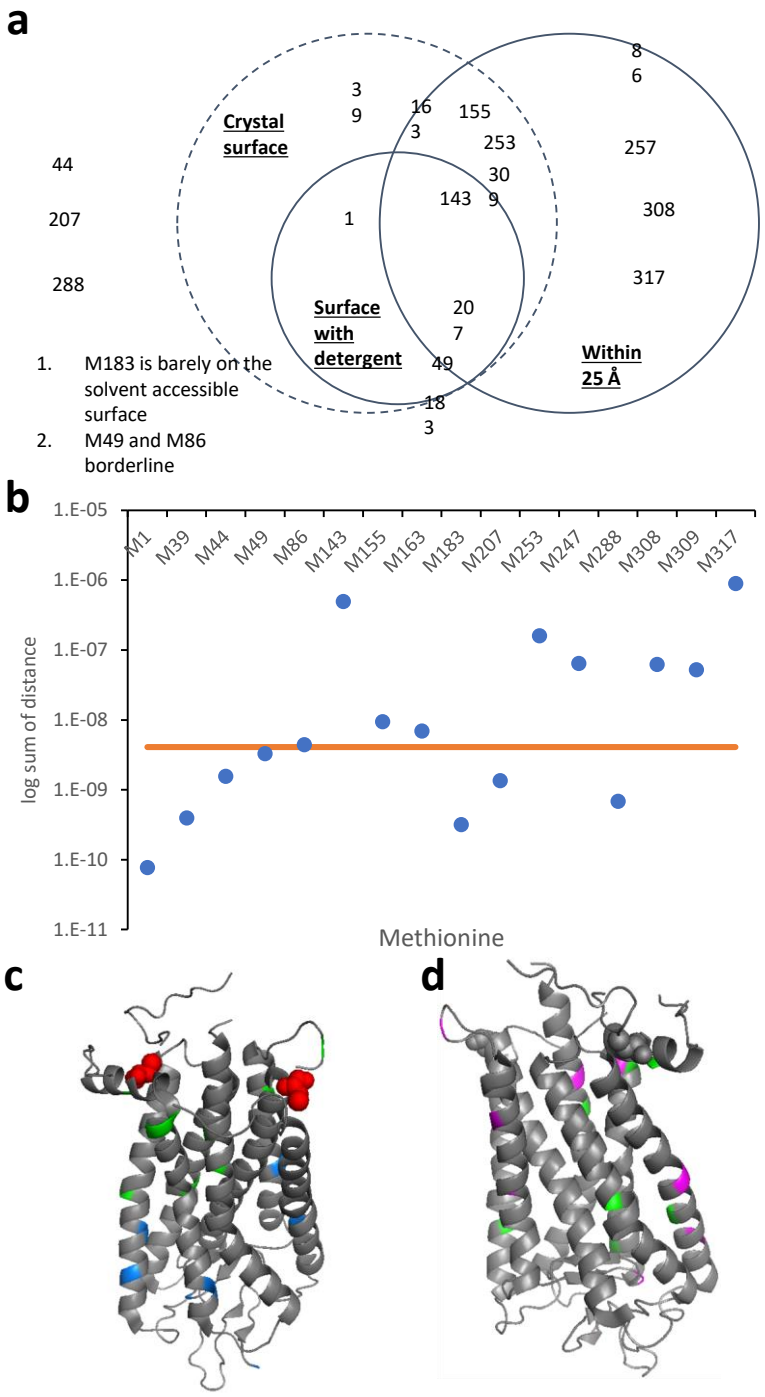


Figure 5.7: Rationale behind MTSL labelling/Soluble PRE reagent. (a) Venn diagram showing the theoretical effects of different PRE reagents on methionines. Residues which are marginal for an effect are placed on the borderlines. **(b)** The sum of the PRE effects on different residues. The closest distances between each methionine and the labelled cysteines was measured (r). The PRE effect is proportional to r^{-6}, and the points represent the sum of these, assuming no synergistic effects. The orange line shows the r^{-6} for 25Å, the quoted value for the limit of PRE effects. The y-axis is logarithmic due to the orders of magnitude difference between residues. Values above the orange line will be affected by MTSL, those below will not. **(c)** Model of rhodopsin displaying labelled cysteines (red spheres), methionines within 25Å of those cysteines (green) and those outside that radius (blue)**. (d)** Model of rhodopsin, displaying methionines on the surface of the crystal structure (magenta), and those buried in the structure (green). Of the nine magenta residues, only four (positions 1, 143, 183, 307) are unlikely to be covered by detergent.

used for EPR spectroscopy (Klein-Seetharaman, 2002) with rhodopsin, and is also, as a spin

label, useful for PRE. Therefore, protocols exist for the MTSL labelling of rhodopsin, though not in the quantities used for NMR. MTSL reacts with accessible cysteines, which in rhodopsin means C140 and C316, both found on the cytosolic side (see figure 5.7c). Past usage of MTSL labelling shows that the biophysical characteristics of rhodopsin are little affected by the label. Existing protocols were modified (see Methods) in an attempt to label only C140. There are also PRE reagents which are soluble, so act upon the residues near the surface.

The rationale for the use of these reagents, alone and in combination, is shown in figure 5.7. The PRE from spin-labels depends on the separation between groups to the power of six, so the magnitude of the effect upon different peaks was used to distinguish peaks near the derivatized residues. The ratio of the effect was estimated by measuring the distance between the thiol group of cysteines 140 and 316 and the methyl groups of each methionine, then adding the two distances to the power of six for each methionine, assuming the effects combine linearly. The added effects are shown in figure 5.7b. The effective range of PRE reagents is usually quoted as 25Å, and 25^{-6} is shown as an orange line. The methionine residues affected by MTSL labelling from these calculations are 86, 143, 155, 163, 247, 253, 308, 309 and 317. M49 is very close to the line with calculations done this way. M183 is not within 25Å of either cysteine, but the combined effects, assuming they are additive, would make it one of the residues affected.

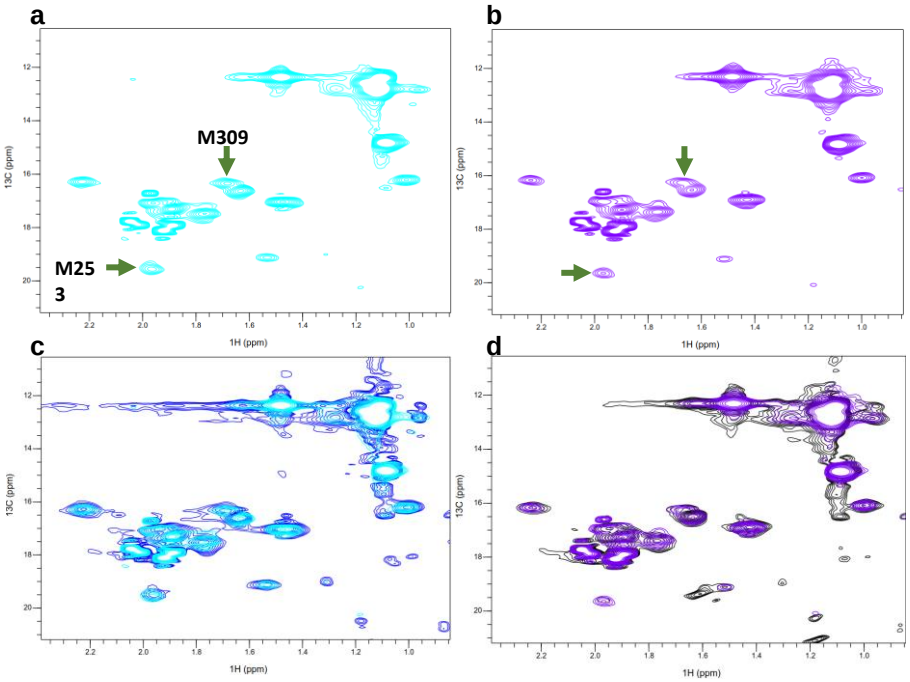


Figure 5.8: Rhodopsin with MTSL labelling (a) MTSL-rhodopsin in DM at 308K **(b)** MTSL-rhodopsin in DM at 298K **(c)** Overlay of (a) with rhodopsin in DM at 308K. **(d)** Overlay of (b) with rhodopsin in DM at 298K.

The extent of labelling achieved by the tested protocols is not entirely certain, though based on previous work with cysteine labelling with PDS (Cai et al., 1999; Judith Klein-Seetharaman et al., 1999) suggest that due to the short time and small concentration chosen, a single cysteine, C316, should be labelled, with possible labelling of a smaller percentage of C140. The positions of peaks in MTSL-rhodopsin are not affected, but some of the relative intensities are (Figure 5.8). The only peak consistently attenuated is the one assigned as M309, which at 298K (Figure 5.8b/d) is at 50% intensity, and at 308K is at 71% intensity (Figure 5.8a/c). The peak assigned as M253 shows 89% intensity at 308K, but 110% intensity at 298K.

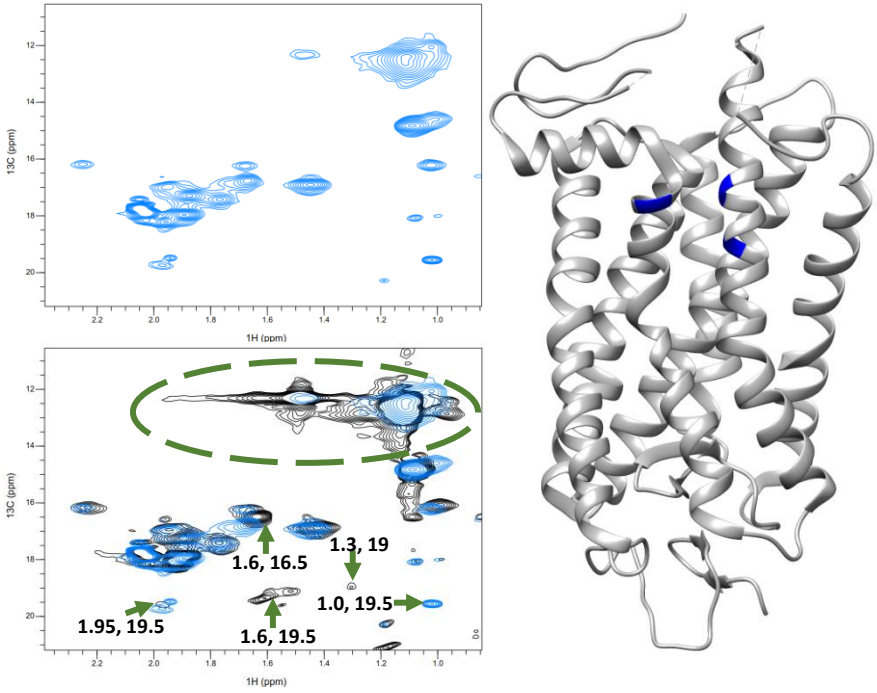


Figure 5.9: HMQC spectrum of rhodopsin in LMNG (a) HMQC spectrum as figure 2e. **(b)** Overlay of figure 2e (cyan, LMNG at 298K) and figure 2d (black, DM at 298K). **(c)** Structure of rhodopsin, with methionines 253, 257 and 308 highlighted in blue.

5.3.4: Rhodopsin in LMNG

DM has strong peaks in NMR, overlapping with the region where methyl resonances are found. Lauryl maltose neopentyl glycol (LMNG) exhibits less interference, and has previously been used for rhodopsin experiments (Gao et al., 2017; Jastrzebska et al., 2013). For these reasons, the possibility of recording NMR spectra in LMNG was explored. A comparison showing the lower interference in LMNG can be seen in figure 5.9b

5.3.4.1: Elution

The solubilisation of rhodopsin by LMNG from ROS suspension was tested, and found not to be sufficiently effective when compared to DM, as shown in figure 5.4b. For this reason,

rhodopsin was solubilised in DM, then exchanged for LMNG during purification as described in section 2.2.1.4.2.2.

When eluting from the 1D4-Sepharose column, 0.001% LMNG in NaPi buffer was first attempted. However, the yield was very low. An increase in LMNG concentration to 0.01% was first attempted, then it was found that switching buffer to PBS, maintaining 0.01% LMNG was necessary. This may increase the portion of misfolded opsin eluted (Liu et al., 1996), but achieved satisfactory elution yields (around 80%).

5.3.4.2: NMR spectra

The HMQC spectrum of rhodopsin in LMNG, shown in Figure 5.9, was not substantially different from that in DM. A large amount of the signal around ^{13}C 12ppm has been lost compared with the DM sample. There are also substantial changes in the spectrum between ^{13}C 19 and 20ppm: the peak(s) around (1.6,19.5) in the DM spectrum have also vanished in the LMNG spectrum. This peak was assigned to M257 in the DM spectrum. The peak at (1.3, 19), assigned to M187, has also disappeared, and an intense peak has appeared at (1.0, 19.5). The peaks assigned to M308 and M253 (at (1.6,16.5) and (1.95,19.5) respectively) have also shifted. This may be of interest, as the residues 308, 253 and 257 are very close. Otherwise, peaks are similar between spectra in the two detergents. Currently, due to the difficulties in solubilisation with LMNG and the overall similarity between the NMR spectra in both detergents, DM is more convenient for NMR experiments.

5.4: Discussion

Introducing ^{13}C into rhodopsin via methionine is useful as a labelling strategy to ease interpretation of NMR spectra of membrane proteins due to methionine's relative scarcity in proteins and slow relaxation characteristics giving relatively sharp lines. Furthermore, focusing on the methyl group in methionine has the advantage that three protons are averaged, multiplying the signal intensity by 3 as compared to a single proton. In rhodopsin, it is also useful because the distribution of methionines is across the protein, including some residues in and near the predicted folding core (figure 1). The assignment strategy using correlations between dihedral angles in crystal structure and chemical shift can be useful but was limited here because the highly dynamic structure of rhodopsin loosened the correlation at the temperatures used for NMR. There are issues with comparing NMR and crystal structures, mainly due to the fact crystallography is conducted ~200K below NMR. This means that there are borderline values for the chemical shifts of methionine for which the conformations are uncertain. Similar difficulties have been encountered in the use of these

strategies with the related β_1AR (Daniel Nietlispach, personal communication). The ^{1}H chemical shifts are very sensitive to ring currents, so any change of the relative positions of these protons to the aromatic residues of rhodopsin will affect peak position.

Expression systems previously used for labelling of other residues have been useful here (Getmanova et al., 2004; J. Klein-Seetharaman et al., 2002; Klein-Seetharaman et al., 2004, 1999; Loewen et al., 2001, p. 19), and have yielded NMR spectra. Though DM has some disadvantages as a detergent for NMR spectroscopy, its utility with rhodopsin has so far outweighed these, though optimisation of elution protocols with LMNG may make this superior in future. It is clear that such improvements are possible as other groups have had greater success (Gao et al., 2017; Jastrzebska et al., 2013). That the amphipol tested here is capable of solubilising rhodopsin is intriguing, due to the potential advantages for NMR of using amphipols over traditional detergents (Etzkorn et al., 2013). The low yield in preliminary tests described here may be improved upon, as amphipols are sensitive to buffer conditions (Popot et al., 2003).

The spatial resolution of the use of PRE reagents may make their use a helpful technique in assignment of spectra, but here the goal has not been fully realised. Very little real change was observed, and the changes disagree with the predicted differences. For example, that the peak assigned to M317 is unaffected is surprising due to its proximity to C316. The peaks that were affected, assigned to M253 and M309, are among those predicted to be strongly affected by MTSL, but the effect upon M253 changes with temperature. Proper determination of the extent of MTSL labelling and optimization of the protocol would improve matters when using the cysteine labelling, and the use of soluble PRE reagents will be more informative than MTSL alone.

CHAPTER 6: RHODOPSIN MISFOLDING

6.1: Summary

The P23H mutation to rhodopsin is a cause of autosomal dominant retinitis pigmentosa (ADRP). The mutant protein can be recombinantly expressed in sufficient quantities for biochemical and biophysical experiments. Previously, western blots qualitatively showed that both P23H and N15S, another ADRP rhodopsin mutant, show aberrant glycosylation. These blots were further analysed to aid interpretation of the qualitative data and supply quantitative analysis. This analysis ascertains that both mutants display different glycosylation from wild-type and that P23H shows more aberrant glycosylation and aggregation than N15S. Samples of methionine labelled P23H in DM were prepared for NMR spectroscopy, building on the work described in chapter 5. Significant differences between wild-type and P23H rhodopsin can be observed in recorded spectra. During attempts to use paramagnetic relaxation enhancement reagents to assign peaks in these spectra, it was found that labelling of P23H cysteines destabilises the protein.

6.2: Introduction

Misfolding of proteins is the underlying cause for a number of diseases (Booth and Curnow, 2006, see section 1.1.2), and a mechanistic understanding of how proteins fold and misfold can have important clinical implications. Furthermore, understanding protein folding is often aided by the study of specific mutations causing misfolding (Dobson, 2003; Fersht, 2000), so efforts to explore membrane protein folding with rhodopsin will benefit from studies of P23H, which misfolds and has compromised thermal stability and light sensitivity (Chen et al., 2014) compared with wild-type. It also displays different glycosylation in Western blots compared to wild-type (Kaushal and Khorana, 1994; Krebs et al., 2010), even when reconstituted with retinal. When reconstituted with 9-*cis*-retinal, the absorbance of the chromophore moves to 476 nm from the more wild-type-like 492 nm displayed when reconstituted with 11-*cis*-retinal (Opefi et al., 2013). Photobleaching and meta II formation are also compromised in reconstituted P23H rhodopsin. Developments of rhodopsin as a folding model, some described in chapter 5, along with the enhancement in P23H expression levels achievable by administrating 9-*cis*-retinal (Noorwez et al., 2004) make it desirable to produce comparable results between wild-type and P23H mutant rhodopsin. It is of further interest that P23 is one of the ADRP mutations found in the predicted folding core of rhodopsin (Klein-Seetharaman, 2005) (see Figure 1.5), as structural and dynamic data for

P23H might show effects consistent with a compromised folding core, and relate these to the other deficiencies in P23H function.

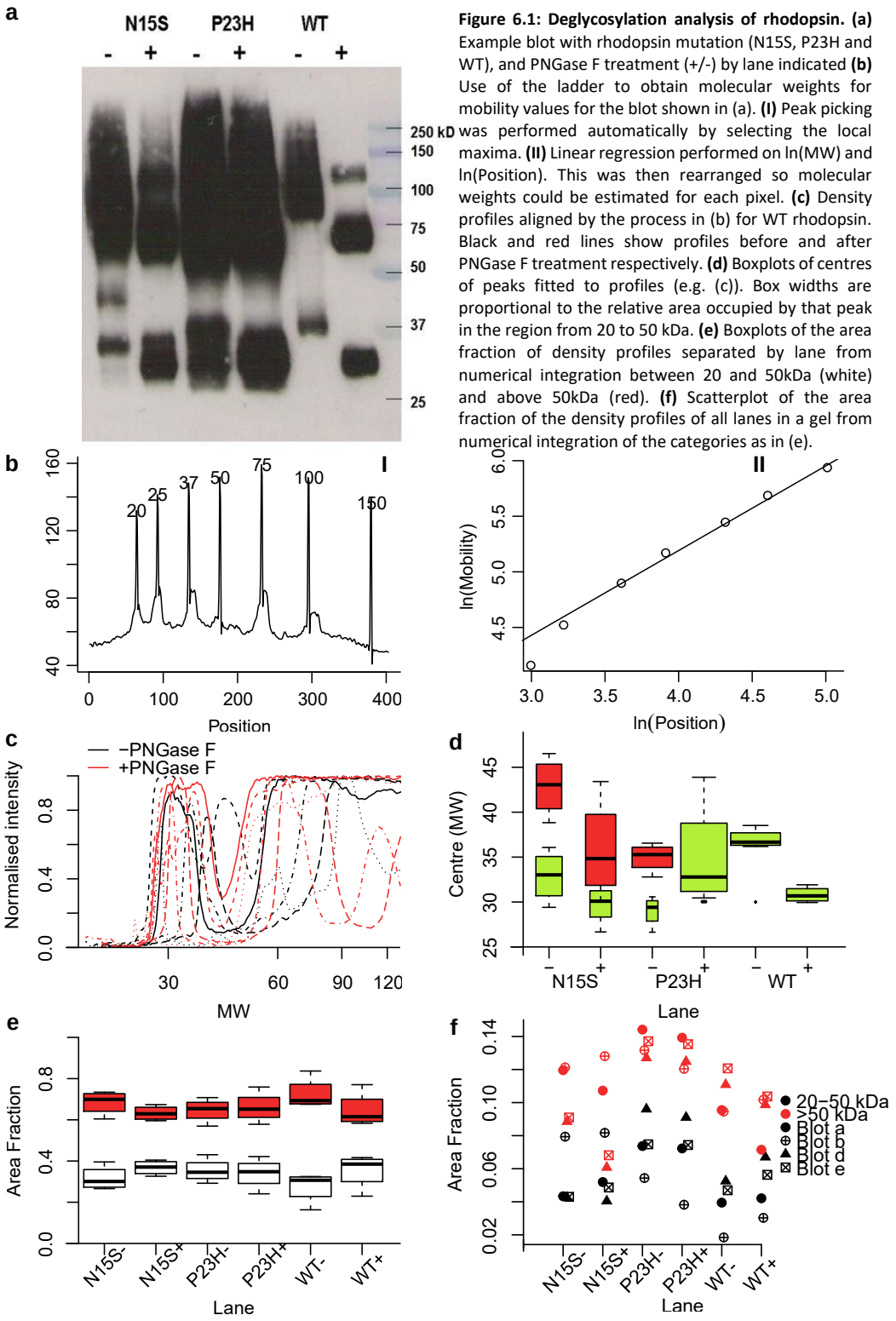


Figure 6.1: Deglycosylation analysis of rhodopsin. (a) Example blot with rhodopsin mutation (N15S, P23H and WT), and PNGase F treatment (+/-) by lane indicated **(b)** Use of the ladder to obtain molecular weights for mobility values for the blot shown in (a). **(I)** Peak picking was performed automatically by selecting the local maxima. **(II)** Linear regression performed on ln(MW) and ln(Position). This was then rearranged so molecular weights could be estimated for each pixel. **(c)** Density profiles aligned by the process in (b) for WT rhodopsin. Black and red lines show profiles before and after PNGase F treatment respectively. **(d)** Boxplots of centres of peaks fitted to profiles (e.g. (c)). Box widths are proportional to the relative area occupied by that peak in the region from 20 to 50 kDa. **(e)** Boxplots of the area fraction of density profiles separated by lane from numerical integration between 20 and 50kDa (white) and above 50kDa (red). **(f)** Scatterplot of the area fraction of the density profiles of all lanes in a gel from numerical integration of the categories as in (e).

6.3: Results

6.3.1: Analysis of deglycosylated rhodopsin immunoblots

Aggregation and glycosylation defects of P23H have been observed (Garriga et al., 1996; Kaushal and Khorana, 1994; Krebs et al., 2010; Tam and Moritz, 2009) (also paper under revision), and to study the phenomena, PNGase F treatment, which removes a variety of N-linked glycosylation (Tarentino et al., 1989), was carried out on P23H and N15S, a mutant affecting one of the glycosylation sites. This treatment removes N-linked glycosylation, and purified naïve and deglycosylated samples of wild-type, P23H, and N15S rhodopsin were resolved by SDS-PAGE and developed by Western blot using the 1D4 antibody raised against the C-terminal nonapeptide of rhodopsin, TETSQVAPA (Molday and MacKenzie, 2002). Previous scans of blots comparing the effects of PNGase F treatment were used (Figure 6.1a). These blots had not been scanned for densitometric analysis, so the scans were saturated, and common analytic techniques could not be applied. They also had no loading controls, so comparison had to be made between relative density of bands. This caused challenges which had to be addressed by the development of a method to obtain information for such saturated scans.

6.3.1.1: Methods development

6.3.1.1.1: Obtaining molecular weights from pixels

Pixel density along a lane can be obtained by using ImageJ (Schneider et al., 2012). This is commonly used for analysis of gels, and has a built-in module for the purpose. However, for comparison between blots, the lane profiles were exported as .csv and loaded into R. The peaks corresponding to ladders were picked as described in Methods, and the natural log pixel position in the profile was fitted against the known molecular weight (MW) of this band by linear regression (Figure 6.1b). From the fitted values, each pixel position could be assigned a molecular weight, which could be compared between different blots (Figure 6.1c).

6.3.1.1.2: Filtering lane profiles

The saturated values of the scans effectively hold no data, so all values at higher than 95% intensity were removed from the data. To subtract the background of each lane, a geometrical “rolling disc” filter was applied, a 2D implementation of the rolling ball filter.

6.3.1.1.3: Fitting bands in lane profiles

To try to recover the data lost by the saturated scans, each band was fitted as a Gaussian peak. For narrow peaks, which had relatively little lost data, the fit of these to the remaining

data was very good, so it was assumed this shape continued into the saturated values. However, for the aggregation bands, the overall distribution was less certain because so many values were missing and the sides of any fitted peak were very steep, implying unrealistically high density values, so this strategy was only employed on bands between 20 and 50 kDa.

The fitted peaks were integrated between 20 and 50 kDa, and their area was calculated as a fraction of the total area of all peaks. The centre of the peak was given as the band's MW and used to construct box and whisker plots (Figure 6.1d). The mean area fraction of each band across the four blots was indicated as the width of its box.

6.3.1.2: Position and density of bands

For WT, there is a single band, found at 37 kDa without PNGase F treatment, and at 31 kDa after.

In P23H, before treatment, one band at 35 and a second, minor band at 29 kDa are observable, and after treatment a single peak at 33 kDa.

N15S has two bands both before and after PNGase F treatment. Before, these are found at 43 and 33 kDa. After, they are found at 35 and 30 kDa.

6.3.1.3: Comparison with aggregated bands

The areas of the lane profiles below and above 50 kDa were calculated using the trapezoidal rule, and the fraction these areas occupied of the total area was compared between lanes (Figure 6.1e,f). After PNGase F treatment, all rhodopsin forms appear to be the same. However, before treatment, both N15S and WT have a greater share of the area in the higher molecular weights.

The increased aggregation found in P23H, even when reconstituted with retinal, suggest associated changes which structural techniques may be able to quantify and use to explain P23H misfolding.

6.3.2: Evidence for structural impairment from Elution Profile

Long incubation times with nonapeptide were required for the elution of WT rhodopsin in LMNG, as shown in section 5.3.4.1, and more misfolded opsin is eluted because of the requirement of using PBS as a buffer instead of NaPi, as can be seen from the higher A_{280} of a rhodopsin sample eluted in PBS in Figure 5.4c. Its compromised thermal stability (Chen et

al., 2014) led to sample loss during the relatively short elution times required when in DM, so to maintain usable yields, P23H was not exchanged to LMNG.

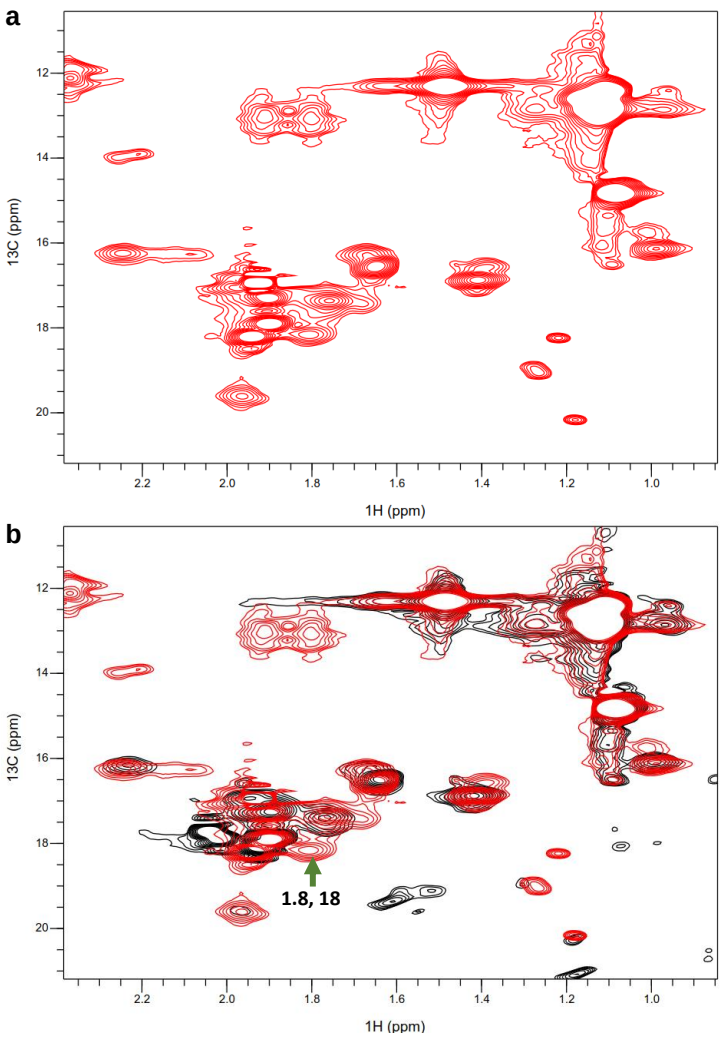


Figure 6.2: HMQC spectra of P23H rhodopsin (a) [$^{13}C_5$, ^{15}N]methionine rhodopsin P23H in DDM, recorded at 298 K with 256 scans at 800 MHz proton frequency. **(b)** Overlay of (a) (red) with [$^{13}C_5$, ^{15}N]methionine rhodopsin (wild type) in DDM, recorded at 298 K with 224 scans at 800 MHz proton frequency.

6.3.3: NMR of methionine labelled P23H rhodopsin

The HSQC spectra of P23H were found to be different from WT spectra recorded under the same conditions (Figure 6.2b). A major difference was the appearance of a strong extra peak at ~(1.8,18)ppm in the P23H spectrum. The position of this peak is within the range of predicted values for the N-terminal methionine (M1). The N-terminal domain is thought to be destabilised by the P23H mutation, leading to the destabilisation of the protein and misfolding, so the appearance of this peak could reflect the difference. The strong signal intensity implies that the N-terminus of rhodopsin has become more flexible in P23H as compared to WT.

There are other, smaller differences in peak positions and relative intensities attesting to the conclusion that long-range perturbations in structure are also present.

To determine whether the strong peak observed in the expected region for M1 in the P23H spectrum, absent from the WT spectrum, was indeed that residue, MTSL spin labels were introduced. The overall rationale for MTSL labelling is described in chapter 5, and in Figure 5.7a it can be seen that M1 could theoretically be uniquely identified using this method.

6.3.4: Cysteine derivatization

MTSL derivatization. Cysteines serve as excellent reporters for structure and dynamics in rhodopsin (Klein-Seetharaman, 2002).The reaction of P23H with two types of labels were investigated. In the first series of experiments, the thiol-specific spin label MTSL, a P23H NMR sample (section 2.2.3.3) was treated with MTSL on the column, but when elution was attempted, yield was negligible.

MTSL has been used with rhodopsin for many EPR experiments (see section 5.3.3), and previous samples of wild-type labelled with MTSL had been eluted successfully, and used for the NMR spectra in Figure 5.8. The loss of yield indicates that MTSL labelling made the P23H protein unstable, most likely through exposure of normally buried cysteines as a result of the P23H mutation.

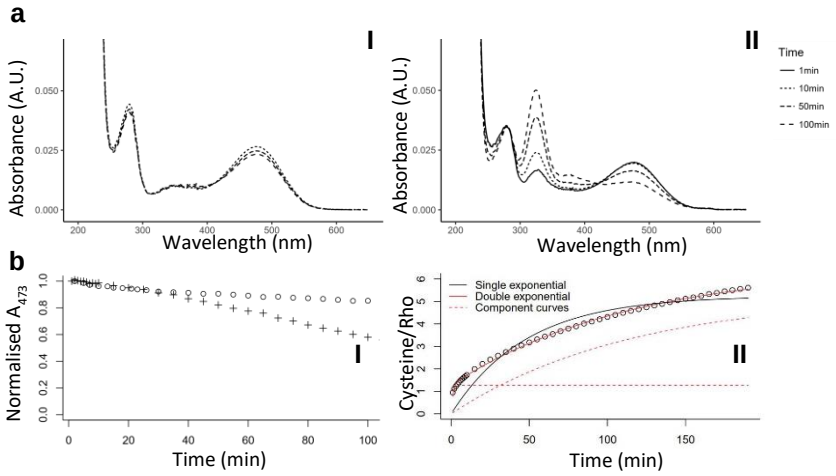


Figure 6.3: Destabilisation of P23H by PDS. (a) Spectra taken from timecourses of 0.5µM P23H in 0.05% DM at 37°C **(I)** in the absence and **(II)** in the presence of 25µM PDS. **(b)** The concentration of P23H relative to t=0 for P23H with (crosses) and without (circles) PDS over time. **(c)** The reaction of PDS with P23H over time. Also shows fitted curves of a single (black) exponential and a double (red) exponential, and the component curves (red dashed line) of the double exponential. The double exponential was preferred by the AIC.

PDS labelling. To further test the hypothesis that cysteine accessibility in P23H rhodopsin was altered, cysteine labelling was carried out using 4-PDS, a thiol label, the binding of which causes an absorbance peak at 323 nm. In WT rhodopsin, two cysteines are accessible (Klein-Seetharaman, 2002), so two PDS molecules react, which is reflected in absorbance spectra.

When conducted on P23H rhodopsin, more than two PDS molecules per rhodopsin reacted. Decomposition of the absorbance spectra suggest all six accessible cysteines reacted (Figure 6.3b II). This indicated that the helical bundle is opened to allow access to PDS to transmembrane helical cysteines. As implied by the MTSL labelling result, derivatization of these cysteines may destabilize P23H rhodopsin. To test this hypothesis, stability of P23H

was measured by losses in chromophore absorbance over time (Figure 6.3b I), and found to be lower in samples with PDS added than in control samples.

Quantifying the rates of reactions are not straightforward because the kinetics of this reaction are complex. This is because we cannot purify PDS-labelled P23H due to the inability to elute from the column, analogous to the MTSL labelling experiment described above. Therefore, chromophore stability was measured in the presence of PDS added. The reaction of PDS with P23H clearly causes destabilisation, which could cause the additional accessibility of cysteines, causing further labelling and possibly increasing destabilisation. Because these interactions could be so non-linear, only rudimentary kinetic analysis was conducted on the time courses obtained. Due to the convenience of PDS as a spectrophotometric measure of cysteine labelling, MTSL labelling was assumed to have the same effect and was not further tested.

6.4: Discussion

6.4.1: Deglycosylation analysis

Analysis of sub-optimal scans of immunoblots has been performed, and show that significant differences in the glycosylation states of rhodopsin mutants can be detected. The differences continue to be observed between the mutants and wild type even after PNGase F treatment. Though the difference between WT and P23H treated with PNGase F is hard to quantify, the persistence of two bands in N15S indicates a difference between two states that is unaffected by PNGase F treatment.

The difference between mutants in the distribution of the higher MW bands could not be well determined with the scans used because of the missing data. For example, from the lane profiles, it looks like P23H has about the same distribution as the others, but from a cursory examination of the blots, it is clear that its high MW bands are much stronger. Unfortunately, due to the missing data in these strong bands, and the relatively strong low MW bands, this cannot be seen in the profile. While reasonable conclusions can be drawn about the distribution of bands below 50 kDa from the existing data, anything more would require new scans tailored for the purpose.

6.4.2: P23H NMR

HMQC spectra of methionine labelled P23H rhodopsin show dramatic differences to WT. Without firm assignments of all peaks, what this shows is not certain, but it demonstrates that methionine is a good residue to label, and that the effects of the mutation are far

reaching. It is possible that, if the extra peak is M1, this residue is missing in the WT construct. However, this is not the case in the original construct and such a spontaneous mutation is unlikely. If necessary, tryptic digests of rhodopsin could be used with mass spectrometry to check the N-terminus. The spectra in DM do have some interference from the DM signal, but the comparison of WT spectra recorded in LMNG suggest that the losses are minimal compared to recording a spectrum in a more NMR-favourable detergent.

While it is unfortunate that MTSL labelling was not viable with P23H rhodopsin, the subsequent PDS experiments were informative. The data fit into an emerging picture of P23H misfolding and abnormal activation behaviour drawn from a combination of molecular dynamics simulations and terahertz spectroscopy recently applied to rhodopsin (Woods, 2014; Woods et al., 2017). From personal communication (Dr K Woods), the P23H mutation introduces strain in the N-terminus disrupting the Cys110-Cys187 disulphide bond. This leads to an influx of water molecules which disrupt hydrogen bonding, leading to misfolding throughout the protein. The disruption by water molecules is also consistent with abnormal activation and Metarhodopsin II formation. Though thiol-specific reagents may not be useful for assignment of P23H peaks, the use of a soluble PRE reagent in future may be helpful for distinguishing surface and buried residues for wild-type.

CHAPTER 7: SUMMARY AND FUTURE WORK

7.1: Summary

Work in this thesis has contributed to our understanding of structure and dynamics in membrane proteins, in particular mammalian rhodopsin and sensory rhodopsin. Understanding of folding and dynamics in membrane proteins lags behind that of soluble proteins (Booth and Curnow, 2006), and developing methods applicable to membrane proteins, establishing model systems for studying folding and dynamics and interpretation of results obtained to date are all important contributions to the field of membrane protein research. This thesis has made strides in each of these directions.

First, to improve our understanding of rhodopsin and vision, I have investigated in detail the interaction of chlorin e6 (Ce6) with rhodopsin and confirmed the enhanced photobleaching of rhodopsin by red light in the presence of Ce6. I have also shown that this effect is dependent on the ratio of Ce6 to rhodopsin. Detailed and deliberate analysis with a low sample requirement has yielded results for the treatment of rhodopsin with an allosteric reagent. I have also shown that both the proposed Ce6 binding site and the G-protein binding domain are well conserved between *Malacosteus niger* and *Bos taurus* rhodopsins, so the observed disruption of G-protein binding to bovine rhodopsin by Ce6 should be maintained in *M niger.* Though it is not as efficient as other reconstitution methods, it is possible to reconstitute rhodopsin with 14-fluororhodopsin in-cuvette. Unfortunately, it is too slow to follow by ^{19}F NMR, but the samples produced are adequate for steady-state spectra, and show no difference in the presence of Ce6. Thus, I have shown it is unlikely that the effects of Ce6 proceed by direct energy transfer, which might be expected to alter the retinal environment.

Second, I have contributed to the establishment of a new model system for membrane protein folding, namely another retinal binding protein, sensory rhodopsin. By modelling the kinetics of pSRII unfolding as measured by different biophysical techniques, I have explored the potential to obtain detailed information about the unfolding reactions from spectroscopic studies. I have written automatic scripts for analysing refolding reactions which have also been used by Yi Lei Tan (University of Cambridge) to analyse their kinetics and efficiency. Adding pSRII to the group of retinal binding proteins which already serve as folding models for membrane proteins further corroborates that proteins with chromophores are particularly useful model systems. Applying principles developed in the

analysis of absorbance spectra, data has been recovered from previous experiments comparing Western blots of rhodopsin with misfolding mutants.

Third, I have progressed the role of rhodopsin in understanding membrane protein folding. NMR experiments have been used to further establish rhodopsin as a model system for class A GPCR folding. The distribution throughout the protein and strong signals given by labelled methionines make this a good starting point for further experiments into rhodopsin folding through examination of unfolded states.

Finally, I have advanced our understanding of rhodopsin misfolding by conducting mechanistic studies of the retinitis pigmentosa mutant, P23H, using NMR spectroscopy of the ^{13}C,^{15}N-methionine labelled protein. The production of useful yields of a misfolding mutant labelled in the same way shows this to be a viable system to explore reasons for folding going wrong as another avenue to understanding the process as a whole. Even without experiments looking at unfolded states, the differences between the spectra of wild-type and mutant rhodopsin support the destabilisation of the previously identified folding core as particularly detrimental to the folding ability and stability of the entire protein. Adding to the NMR experiments, I made a serendipitous discovery that cysteine labelling further destabilises P23H, a result consistent with predictions made from molecular dynamics.

These contributions have opened doors to a number of future studies, described in detail below. They are organized by immediate experiments that logically follow from the work of this thesis, and by long-term studies that will significantly advance the field of membrane protein research.

7.2: Future Work

7.2.1: Immediate goals of future work

7.2.1.2: Rhodopsin and vision

7.2.1.2.1: Enhancement of bleaching by Chlorin e6

That Ce6 enhances the bleaching of rhodopsin with light in the far red has been demonstrated in previous work. It also appears that this is somewhat specific, as the dark noise of rhodopsin is not similarly enhanced, as shown in chapter 3. Further specifying the light-dependence of this enhancement might further illuminate the mechanism. If the

enhancement is wavelength dependent, the proposed triplet energy transfer mechanism would be further supported. More wavelength specific filters, including bandpass filters, are available and could be used to narrow down the wavelengths at which the greatest enhancement takes place. Another possibility would be to use lasers, which can be tuned to specific wavelengths, and might be more cost-efficient than specific filters.

Another possibility would be to use neutral density filters and the standard yellow filter used for rhodopsin bleaching to determine whether a more general bleaching enhancement occurs. For any of this to be informative, a more reproducible protocol needs to be developed. Though a general increase of the enhancement is observed with a higher Ce6:rhodopsin ratio, the rates of bleaching are highly variable, especially at higher Ce6. Refinements in the illumination procedure and the reliable addition of quantities of Ce6 might improve reproducibility. Once this is achieved, given the low affinity of Ce6 for rhodopsin, higher Ce6:rhodopsin ratios might be tested to find the limits of bleaching rate enhancement.

It has been shown that Ce6 enhances rhodopsin stability (Fernanda Balem et al., 2009), and also tentatively shown that it at least does not increase thermal activation. The data show some reduction in thermal activation, but have not been replicated. Repeats should be performed to measure the effect at different temperatures.

For Ce6 to enhance rhodopsin's bleaching rate, some form of association between the two molecules must occur. The nature of this association is being explored through THz spectroscopy (Woods et al., 2017), but absorbance and phosphorescence spectroscopy may also have additional information to offer. Past fluorescence spectroscopy has not shown energy transfer (Ilyas Washington et al., 2004), which has been interpreted as lack of singlet energy transfer. Instead, a triplet energy transfer model has been proposed but not tested. Intersystem crossing to enable triplet energy transfer would provide another path for triplet states to follow and cause quenching of Ce6 phosphorescence upon addition of rhodopsin, and characterising this interaction would put some constraints on the nature of the mechanism at work. Such experiments require the detection of chlorin phosphorescence. The calculated emission spectrum for Ce6 phosphorescence has a λ_{max} of 1080nm (Gattuso et al., 2017).

A characteristic of porphyrins is their ability to bind metals in the macrocycle. Rhodopsin binds several zinc atoms. Though no obvious changes indicating zinc binding were observed in the recorded spectra, rhodopsin samples were prepared with EDTA during the Ce6

experiments and it may be worth investigating how zinc free these samples are and whether this affects the effects of Ce6. On one hand, the changes metal ion binding causes to chlorin spectra are dramatic; the Q-band would be reduced to two peaks rather than four. On the other, Ce6's affinity for rhodopsin is low, so with the high ratios of chlorin to rhodopsin used only a minority of Ce6 molecules would be bound to rhodopsin. Alternatives might be to add zinc and/or EDTA to rhodopsin/Ce6 mixtures and measure bleaching rates. From trying to observe the effects of chlorin upon SDS denaturation of rhodopsin it was observed that Ce6 absorbance spectra are sensitive to SDS concentration, at least in the range from 0 to 3% SDS. Refinements in fitting methods applied to Ce6 spectra might be able to isolate different changes in Ce6 spectra with better sensitivity, including some minor effects that might be discernible in properly decomposed spectra of rhodopsin/Ce6 mixtures. The transitions that lead to the different absorbance bands of porphyrins are well understood, and detailed information of this kind might be applied to any changes in chlorin spectra so absorbance spectroscopy could provide insight into Ce6 binding to rhodopsin.

7.2.1.2.2: NMR of Rhodopsin and Ce6

The in-cuvette reconstitution of rhodopsin with 14-fluororetinal was only 43% efficient. Though the addition of Ce6 does not appear to have any effect upon its NMR spectrum, for any other experiments on this system, a better sample may offer greater sensitivity. Existing reconstitution protocols have higher yields of fluororhodopsin (Steinberg et al., 1993), and as the kinetics of reconstitution are not accessible to a ^{19}F NMR experiment, these could be used instead. The current samples also have a background of excess 14-fluororetinal, which would not be an issue with other preparation methods. Binding reconstituted rhodopsin onto a 1D4 column and washing away excess retinal will address this issue. The comparison of fluororhodopsin with the fluororetinal used to reconstitute it shows changes in peaks associated with the free pigment of other fluorine analogues, a result which will need explanation.

^{15}N labelling lysines in rhodopsin has been used in the past for NMR studies highlighting the cytoplasmic domain (J. Klein-Seetharaman et al., 2002). This could also be used to test the predicted cytoplasmic Ce6 binding site. K339 in particular is very flexible, showing a signal like that of lysine in solution. If Ce6 binds the C-terminal domain (including the seven predicted residues missing from *M. niger's* partial sequence) it is possible this flexibility would be reduced.

7.2.1.3: Establishment of pSRII as an experimental model system for unfolding

7.2.1.3.1: Biophysical kinetic studies of pSRII

Analysis of pSRII spectra has established a sensible model for its unfolding with SDS, but the experimental setup used has limited this to relatively late events. The first spectra were recorded after a minute or more, and already represent a significant departure from a native spectrum. New models are likely to be necessary for rapid-mixing experiments proposed by our Cambridge collaborators, especially for the different kinetics that may be observed for the disappearance of shoulder peaks in the absorbance spectra. It will also be interesting to establish whether the transition between the native maximum absorbance and the red-shifted species observed can be captured with early spectra.

The ambiguity in hydroxylamine fluorescence time courses suggests that two processes with similar rates are occurring, the loosening of the helical bundle, and the formation of retinal oxime. The second rate may be close enough to the first that the two curves cannot be separated effectively by model selection methods. If the unreliable fitting of the time courses truly is physical rather than artifactual, hydroxylamine may be useful for further exploring the amplitude of the fluorescence curves as sources of information about the distances between retinal and pSRII's tryptophans, and thus, the nature of the loosened helical bundle caused by SDS unfolding.

For expediency and ease of fitting, the model for fitting of fluorescence time courses used omitted the contribution of native pSRII. It is possible that this, while useful at the time, caused loss of information that could be valuable. The kinetic model finally used for the fitting of absorbance spectra introduced an error term for the dead time of the experiment. A model including this might be able to fit kinetics curves for all three species. It is not certain this will yield any new information from existing time courses, but given the dominance of native pSRII at early time points, it may be useful for stopped-flow experiments.

Automatic peak picking

With the forgiving Levenberg-Marquardt algorithm for non-linear least squares, fairly poor initial estimates often find a good fit. The estimates provided by the peak picking algorithm are close to those yielded by the final model and little optimisation is required. It is unfortunate that the underlying assumption of symmetrical peaks was not realistic. The peak picking performed well on the data when the maximum number of peaks was externally defined, but once model selection for the number of peaks was implemented, the skewedness of the shoulder peaks distorted the output, adding a peak to the fit. It is possible

that a different model selection method, such as cross-validation, would not be “fooled” by this, though this was not observed when comparing choices between fluorescence time course models. Later fits were improved by the addition of a flexible scattering baseline, which might prove useful in predicting peaks.

The skew observed in pSRII is not obvious by visual inspection and the recognition of such features by computer may not be simple either. The exhaustive solution would be to add skew to each peak in turn and test whether it improves the model, but as skewed peaks may not be common, the added processing time might limit the utility of such an approach. It is also likely that, in a spectrum with multiple skewed peaks, the skews might interact in unpredictable ways so that a single skewed peak may not improve a fit, but two or more might. Fitting this behaviour would require adding skew to all combinations of peaks. As this algorithm is only likely to be useful with many, overlapping peaks, processing times would become untenable.

Even the simple clustering approach used here identified reasonable patterns in the peaks, especially given the variability caused by the unrealistic model. More advanced techniques are far more capable, so the refinement of a single time point by using the entire dataset is a promising avenue for improving prediction. The real promise of such a method is that it might make the recognition of kinetics in the species present possible. For example, any peaks that have a statistically significant change over time could be made candidates for kinetic analysis. Any peaks in a constant ratio could be treated as one species, then guesses as to the kinetics of remaining candidate species could be made, given the number of species present and known rate equations. Ideally, time courses of spectra which a user suspects display some unknown kinetics could be used as input, and parameters for a kinetic model like that derived for pSRII unfolding in SDS would be the output. With the limited success of the relatively undeveloped algorithm described, this does not seem unrealistic, given the low-hanging fruit of the optimisation candidates described above.

7.2.1.3.2: Analysis of pSRII NMR spectra

The original goal of using PCA on pSRII was to impute values for missing peaks. However, the sparse spectra at higher concentrations of SDS limit this goal, and the greatest amount of imputation that can be relied upon has probably been achieved, as the expectation maximization algorithm used performs robustly compared with other methods. The fact two principal components were sufficient to explain so much of the variation, and that the scores for these principal components follow a definite pattern, suggest that the PCA has identified

underlying physical processes. The regularity also displayed by the loadings for each residue when a polar plot was used further suggest that these processes might be amenable to further human inspection.

A major shortcoming of the PCA analysis is that the ^{1}H and ^{15}N chemical shifts of a residue are separated, so significant information is lost. There are analogous methods to PCA for 3-dimensional arrays, where PCA operates on a 2-dimensional matrix. This would allow association of the chemical shifts for each residue by having one dimension as [SDS], one as the residue number, and one as the atom. This would also allow proper incorporation of peak intensity data also. In this analysis, peak intensity was discarded because of the difficulty of scaling this data with chemical shifts, but working with the higher dimensional methods may mitigate some of the issues of poor scaling encountered with PCA. Attempts were made to use the Tucker3 algorithm (Andersson and Bro, 1998) on the chemical shift dataset, but in the implementation used there is no imputation process, so the dataset was further limited. In future, either the dataset as imputed by pPCA will be used with the Tucker3 algorithm, or another implementation will be used. There are two readily available R packages that may be suitable, PTAk and rTensor, or the N-way toolbox for MATLAB, which includes an Expectation Maximization algorithm for imputation.

Clustering the profiles of chemical shift change against SDS concentration showed significant dependence of these profiles on the domain of the residue. This was relatively crude, as it only looked at the magnitude of the change in the ^{13}C and a scaled ^{15}N dimension. Use of clustering methods applied to a more sophisticated representation of the data which allows both positive and negative changes in multiple dimensions may be even more sensitive and provide insight as to the physical nature of the grouping of residues.

Ultimately, the aim of blob analysis, which is a simple algorithm for recognising objects in images, was to perform it on all spectra, then try to identify similar blobs across spectra to track the motion of peaks. There are techniques used in computer vision to track and predict future positions of some feature. The Kalman filter is relatively simple to implement, and performs well in this regard. However, it relies on iteration, so the small number of measurements in the NMR dataset may be too few for this to be effective. By using some feature recognition and prediction algorithms to weight regions of spectra, it may be possible to bring out peaks from noise without relying on some model for their movement. In the future, feature detection performed on higher resolution spectra may be useful for recovering data when there is good reason to suppose a regular pattern between spectra.

7.2.1.4: Further Development of Rhodopsin as a model system for NMR studies of folding and misfolding

Methionine labelling of rhodopsin shows great promise as a method for probing its folding core by NMR spectroscopy. The assignments made from the theoretical calculations based on torsion angles are consistent with expectations of flexibility, though they are held to be ambiguous between chemical shifts of 16.5 and 18. Practical problems with the sample preparation of MTSL labelled samples have prevented this technique from making more definitive assignments, so the use of a more optimised protocol might allow more use to be made of the HMQC spectra already acquired, as described in the original plan for the MTSL spectra. A further possibility using current samples would be to attempt assignment of M308 and M309 using a through-bond HNCO experiment. However, HNCO experiments have been challenging with rhodopsin in the past and generally only work for highly flexible residues since they require use of the backbone ^{15}N resonance(J. Klein-Seetharaman et al., 2002). Mutagenesis has been required in past assignment strategies (Werner et al., 2008) and the introduction of mutants can cause distortions of the protein, resulting in unwanted chemical shift changes. Surface residues can be separated from buried residues by H/D exchange and PRE experiments with MTSL labels can be carried out for WT, where MTSL label does not interfere with protein stability.

Once definitive assignments have been made, NMR relaxation experiments can be carried out to assess the flexibility of the different regions. Unfolding experiments in SDS can take advantage of these experiments and the determination of ideal unfolding conditions for rhodopsin previously established.

7.2.1.5: Misfolding of rhodopsin

7.2.1.5.1: P23H NMR spectra

The first priority in the improvement of P23H NMR data is to improve the yields of pure methionine labelled P23H samples. Similar yields of P23H were found in cell lysate when retinal was administered to spinner flasks as were found in wild type samples. The difference in elution of P23H from 1D4 columns using PBS elution buffer rather than sodium phosphate suggest significant improvements might be made by using PBS buffers straight away. This was attempted when eluting MTSL labelled sample, but the destabilising effect of cysteine labelling meant that this could not serve as a test of using that protocol to improve yields.

The differences shown between the HMQC spectra of WT and P23H rhodopsin show promise for demonstrating the importance of residues in the proposed folding core. However, the

assignment of the massively increased peak as M1 should be confirmed by the use of MTSL labelling and the use of a soluble PRE reagent. This cannot be demonstrated directly, as the labelling of cysteines causes catastrophic degradation of P23H. Instead, the comparison with properly MTSL assigned WT spectra must be made. It is possible that some confirmation can be made using a soluble PRE reagent.

7.2.1.5.2: Cysteine labelling

That PDS labelling, usually benign for WT rhodopsin, should destabilise P23H, is very interesting. The methods used to fit the spectra required the fixing of some parameters in the model after obtaining estimates. This was avoided in the fitting of SDS unfolding time courses by using a combined kinetics model, allowing the use of the whole dataset to estimate the parameters for individual peaks. Ideally, this would be done for PDS unfolding too, but the kinetics are very complex and non-linear, as described in the discussion section of that chapter. One possibility for simplifying the kinetics would be by mutagenesis, removing cysteines from the protein, which would also determine which of the cysteines destabilise the protein after labelling. This would be a very effortful approach, and the rewards may not justify the effort. One lower-effort improvement to the experiment might be in the analysis. Non-linear mixed-effects models (NLME) allow different levels of fitting for different parameters when fitting a multivariate dataset. Using an NLME model, the widths and positions of peaks could be held to be constant across time points, with variation due to random error, while the amplitudes could be allowed to vary freely over time. This would allow the use of the whole dataset to most accurately estimate the peaks' shapes without an explicit kinetic model. Another refinement of the experiment is to choose some time points at which to stop the reaction by adding cysteine to the cuvette. By using several of these time points, the kinetics of the labelling of whichever cysteine causes destabilisation could be estimated by the extent of unfolding caused at different times of arresting the labelling. This could be compared with the kinetics of labelling to determine how quickly the labelling occurs to give clues as to which residue(s) are important.

It is also important that the constructs used are checked so it is certain that the wild type construct includes an N-terminal methionine. The possible digests of rhodopsin are well chararacterised, so mass spectrometry can be used to confirm this.

7.2.2: Long term goals of future work

7.2.2.1: Rhodopsin and vision

The enhancement of rhodopsin bleaching by red light upon addition of Ce6 has been demonstrated both in previous literature, and more definitively here. That this enhancement *in vitro* causes a similar enhancement of vision *in vivo* is less certain. There have been demonstrations in mouse models of an increased retina response to both blue and red light with addition of Ce6, though a G-protein binding assay shows activation is inhibited by Ce6 (paper in preparation), so the role of chlorin is still uncertain. If it works as a red filter, rather than a simple signal booster, as proposed in the discussion of the Rhodopsin and Vision chapter, the effect might be challenging to demonstrate. However, if the thermal activation reduction hinted at by preliminary results can be better characterised, and the bleaching rate enhancement more reliably demonstrated, computational models accounting for these effects and the inhibition of G-protein activation might be devised to show that increasing the sensitivity of cells to the activation of rhodopsin molecules, as suggested for the use of tapeta in eyes with less pigment, would be able to better detect the reflection of red bioluminescence.

What else chlorin might enhance might give clues as to the nature of the chlorin-rhodopsin interaction. If chlorin binds other GPCRs, that would be interesting because it shows Ce6 to be a promiscuous partner for that whole family. If chlorin enhances the activity of other retinal binding proteins, that shows the interaction is not protein specific, and suggests the bleaching enhancement is simply due to general energy transfer. If the effect is present, but different, for retinal binding proteins, the differences between proteins might give clues to the nature of the interaction. If neither case can be shown, despite the rhodopsin bleaching enhancement across such an evolutionary distance as between *M. niger* and *B. taurus*, that would be remarkable and represent some unappreciated role of chlorophyll derivatives in the evolution of animal vision.

7.2.2.2: pSRII as a folding model

The biophysical characterisation of the unfolding kinetics of pSRII in SDS was useful for establishing the system as an experimental model. However, these do not directly address the question of similarities and differences between MR, pSRII, and bR. However, the various analytical techniques used to explore the rich NMR dataset may yet distinguish between local and long-range interactions in the unfolded states. Exploring the loadings of different residues in the PCA may be aided by comparison with improved clustering approaches to

chemical shift change data. The incorporation of peak intensity data will provide information about the flexibility of different regions and may provide further principal components which can be associated with particular regions in unfolding and checked against FIRST analysis results for pSRII.

7.2.2.3: Rhodopsin as a folding model

Long term, the long-range interactions model for rhodopsin needs to be demonstrated by testing specific predictions made against the two-stage model. With even partial assignments of the methionine spectra, differences caused in unfolding by SDS may be amenable to similar analysis carried out on pSRII.

The use of PRE reagents will be informative for the reliable assignment of peaks, which will be helpful to both further NMR experiments into rhodopsin unfolding and potentially the NMR experiments of other investigators, as methionine's characteristics make it a useful residue to label. Use of MTSL and soluble PRE reagents might help directly with interpretation of NMR spectra of unfolding rhodopsin in other ways. If the soluble reagents show solvent accessibility in some way, the extent of unfolding might be detectable, and possibly even kinetics if the acquisition can be fast enough. If some peaks that disappear after proper MTSL preparation reappear after unfolding, it might show that the methionines get further away from the MTSL labelled cysteines. Those are not on the folding core side, so the SDS dependence of any effect like this might be informative.

7.2.2.4: Misfolding of rhodopsin

Even the comparisons made between native WT and P23H NMR spectra in this thesis have shown differences. The possibilities discussed above for NMR spectroscopy of WT rhodopsin could also be contrasted with the same experiments carried out on P23H. For example, any changes shown in unfolding WT might be comparable to native P23H, or happen at lower SDS in P23H.

The destabilisation of P23H by thiol-specific reagents may also be of interest in future experiments. The central role of the conserved disulphide bond to the stability of rhodopsin is well recorded, and there are preliminary results from molecular dynamics simulations that this bond is somewhat compromised in P23H (Dr Kristina Woods, personal communications). It may be that cysteine labelling could offer productive avenues to the study of misfolding in P23H and other rhodopsin mutants, especially those associated with retinitis pigmentosa.

BIBLIOGRAPHY

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alexiev, U., Farrens, D.L., 2014. Fluorescence spectroscopy of rhodopsins: Insights and approaches. Biochim. Biophys. Acta BBA - Bioenerg. 1837, 694–709. https://doi.org/10.1016/j.bbabio.2013.10.008

Andersson, C.A., Bro, R., 1998. Improving the speed of multi-way algorithms:: Part I. Tucker3. Chemom. Intell. Lab. Syst. 42, 93–103. https://doi.org/10.1016/S0169-7439(98)00010-0

Applebury, M.L., Hargrave, P.A., 1986. Molecular biology of the visual pigments. Vision Res. 26, 1881–1895.

Arst, H., Mäekivi, S., Lukk, T., Herlevi, A., 1997. Calculating irradiance penetration into water bodies from the measured beam attenuation coefficient. Limnol. Oceanogr. 42, 379–385. https://doi.org/10.4319/lo.1997.42.2.0379

Athanasiou, D., Aguila, M., Bellingham, J., Li, W., McCulley, C., Reeves, P.J., Cheetham, M.E., 2018. The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. Prog. Retin. Eye Res. 62, 1–23. https://doi.org/10.1016/j.preteyeres.2017.10.002

Atkins, P., Atkins, P.W., Paula, J. de, 2014. Atkins' Physical Chemistry. OUP Oxford.

Aue, W.P., Bartholdi, E., Ernst, R.R., 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J. Chem. Phys. 64, 2229–2246. https://doi.org/10.1063/1.432450

Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., Lanyi, J.K., 2005. Xanthorhodopsin: A Proton Pump with a Light-Harvesting Carotenoid Antenna. Science 309, 2061–2064.

Balem, F., Yanamala, N., Klein-Seetharaman, J., 2009. Additive effects of chlorin e6 and metal ion binding on the thermal stability of rhodopsin in vitro. Photochem Photobiol 85, 471–8. https://doi.org/10.1111/j.1751-1097.2009.00539.x

Balem, Fernanda, Yanamala, N., Klein-Seetharaman, J., 2009. Additive Effects of Chlorin E6 and Metal Ion Binding on the Thermal Stability of Rhodopsin In Vitro†. Photochem. Photobiol. 85, 471–478. https://doi.org/10.1111/j.1751-1097.2009.00539.x

Barker, B.E., Fox, M.F., 1980. Computer resolution of overlapping electronic absorption bands. Chem. Soc. Rev. 9, 143. https://doi.org/10.1039/cs9800900143

Bax, A., Griffey, R.H., Hawkins, B.L., 1983. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. J. Magn. Reson. 1969 55, 301–315. https://doi.org/10.1016/0022-2364(83)90241-X

Bayburt, T.H., Sligar, S.G., 2010. Membrane protein assembly into Nanodiscs. FEBS Lett 584, 1721–7. https://doi.org/10.1016/j.febslet.2009.10.024

Baylor, D.A., Lamb, T.D., Yau, K.W., 1979. Responses of retinal rods to single photons. J. Physiol. 288, 613–634.

Bazzacco, P., Sharma, K.S., Durand, G., Giusti, F., Ebel, C., Popot, J.-L., Pucci, B., 2009. Trapping and Stabilization of Integral Membrane Proteins by Hydrophobically Grafted Glucose-Based Telomers. Biomacromolecules 10, 3317–3326. https://doi.org/10.1021/bm900938w

Beatty, E.J., Cox, M.C., Frenkiel, T.A., Tam, B.M., Mason, A.B., MacGillivray, R.T.A., Sadler, P.J., Woodworth, R.C., 1996. Interlobe Communication in ^{13}C-Methionine-Labeled Human Transferrin †. Biochemistry (Mosc.) 35, 7635–7642. https://doi.org/10.1021/bi960684g

Beecham Matthew P., Clarkson Guy J., Hall Gareth, Marsh Andrew, 2013. Nanostructures from Self-Assembling Triazine Tertiary Amine N-Oxide Amphiphiles. ChemPhysChem 14, 3909–3915. https://doi.org/10.1002/cphc.201300775

Berson, E.L., Rosner, B., Weigel-DiFranco, C., Dryja, T.P., Sandberg, M.A., 2002. Disease Progression in Patients with Dominant Retinitis Pigmentosa and Rhodopsin Mutations. Invest. Ophthalmol. Vis. Sci. 43, 3027–3036.

Bodenhausen, G., Ruben, D.J., 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett. 69, 185–189. https://doi.org/10.1016/0009-2614(80)80041-8

Booth, P., Curnow, P., 2006. Membrane proteins shape up: understanding in vitro folding. Curr. Opin. Struct. Biol. 16, 480–488. https://doi.org/10.1016/j.sbi.2006.06.004

Bowmaker, J.K., 1995. The visual pigments of fish. Prog. Retin. Eye Res. 15, 1–31. https://doi.org/10.1016/1350-9462(95)00001-1

Bowmaker, J.K., Dartnall, H.J.A., Herring, P.J., 1988. Longwave-sensitive visual pigments in some deep-sea fishes: segregation of 'paired' rhodopsins and porphyropsins. J. Comp. Physiol. A 163, 685–698. https://doi.org/10.1007/BF00603853

Bracken, C., 2001. NMR spin relaxation methods for characterization of disorder and folding in proteins. J. Mol. Graph. Model. 19, 3–12. https://doi.org/10.1016/S1093-3263(00)00136-4

Bravaya, K., Bochenkova, A., Granovsky, A., Nemukhin, A., 2007. An Opsin Shift in Rhodopsin: Retinal S0−S1 Excitation in Protein, in Solution, and in the Gas Phase. J. Am. Chem. Soc. 129, 13035–13042. https://doi.org/10.1021/ja0732126

Butterfoss, G.L., DeRose, E.F., Gabel, S.A., Perera, L., Krahn, J.M., Mueller, G.A., Zheng, X., London, R.E., 2010. Conformational dependence of 13C shielding and coupling constants for methionine methyl groups. J. Biomol. NMR 48, 31–47. https://doi.org/10.1007/s10858-010-9436-6

Cai, K., Klein-Seetharaman, J., Farrens, D., Zhang, C., Altenbach, C., Hubbell, W.L., Khorana, H.G., 1999. Single-Cysteine Substitution Mutants at Amino Acid Positions 306−321 in Rhodopsin, the Sequence between the Cytoplasmic End of Helix VII and the Palmitoylation Sites: Sulfhydryl Reactivity and Transducin Activation Reveal a Tertiary Structure,. Biochemistry (Mosc.) 38, 7925–7930. https://doi.org/10.1021/bi9900119

Calabrese, A.N., Watkinson, T.G., Henderson, P.J.F., Radford, S.E., Ashcroft, A.E., 2015. Amphipols Outperform Dodecylmaltoside Micelles in Stabilizing Membrane Protein Structure in the Gas Phase. Anal. Chem. 87, 1118–1126. https://doi.org/10.1021/ac5037022

Campbell, A.K., Herring, P.J., 1987. A novel red fluorescent protein from the deep sea luminous fish Malacosteus niger. Comp. Biochem. Physiol. Part B Comp. Biochem. 86, 411–417. https://doi.org/10.1016/0305-0491(87)90314-2

Cangelosi, R., Goriely, A., 2007. Component retention in principal component analysis with application to cDNA microarray data. Biol. Direct 2, 2. https://doi.org/10.1186/1745-6150-2-2

Carpenter, E.P., Beis, K., Cameron, A.D., Iwata, S., 2008. Overcoming the challenges of membrane protein crystallography. Curr. Opin. Struct. Biol. 18, 581–586. https://doi.org/10.1016/j.sbi.2008.07.001

Chen, Y., Jastrzebska, B., Cao, P., Zhang, J., Wang, B., Sun, W., Yuan, Y., Feng, Z., Palczewski, K., 2014. Inherent Instability of the Retinitis Pigmentosa P23H Mutant Opsin. J. Biol. Chem. 289, 9288–9303. https://doi.org/10.1074/jbc.M114.551713

Choe, H.-W., Kim, Y.J., Park, J.H., Morizumi, T., Pai, E.F., Krauß, N., Hofmann, K.P., Scheerer, P., Ernst, O.P., 2011. Crystal structure of metarhodopsin II. Nature 471, 651. https://doi.org/10.1038/nature09789

Chung, K.Y., Kim, T.H., Manglik, A., Alvares, R., Kobilka, B.K., Prosser, R.S., 2012. Role of Detergents in Conformational Exchange of a G Protein-coupled Receptor. J. Biol. Chem. 287, 36305–36311. https://doi.org/10.1074/jbc.M112.406371

Cideciyan, A.V., Hood, D.C., Huang, Y., Banin, E., Li, Z.-Y., Stone, E.M., Milam, A.H., Jacobson, S.G., 1998. Disease sequence from mutant rhodopsin allele to rod and cone photoreceptor degeneration in man. Proc. Natl. Acad. Sci. 95, 7103–7108.

Colmenares, L., 1991. NMR studies of fluorinated visual pigment analogs. Biochem. Biophys. Res. Commun. 179. https://doi.org/10.1016/0006-291X(91)91720-W

Curnow, P., Booth, P.J., 2007. Combined kinetic and thermodynamic analysis of α-helical membrane protein unfolding. PNAS. https://doi.org/10.1073/pnas.0705067104

Dahmane, T., Damian, M., Mary, S., Popot, J.-L., Banères, J.-L., 2009. Amphipol-Assisted in Vitro Folding of G Protein-Coupled Receptors. Biochemistry (Mosc.) 48, 6516–6521. https://doi.org/10.1021/bi801729z

Daiger, S., Sullivan, L., Bowne, S., 2013. Genes and mutations causing retinitis pigmentosa. Clin. Genet. 84. https://doi.org/10.1111/cge.12203

Davie, J.R., 2003. Inhibition of Histone Deacetylase Activity by Butyrate. J. Nutr. 133, 2485S-2493S. https://doi.org/10.1093/jn/133.7.2485S

De Bernardez Clark, E., Schwarz, E., Rudolph, R., 1999. Inhibition of aggregation side reactions during in vitro protein folding. Methods Enzymol. 309, 217–236.

Denton, E.J., Nicol, J. a. C., 1964. The Chorioidal Tapeta of some Cartilaginous Fishes (Chondrichthyes). J. Mar. Biol. Assoc. U. K. 44, 219–258. https://doi.org/10.1017/S0025315400024760

Dexter, D.L., 1953. A Theory of Sensitized Luminescence in Solids. J. Chem. Phys. 21, 836–850. https://doi.org/10.1063/1.1699044

Dobson, C.M., 2003. Protein folding and misfolding 426, 7.

Douglas, R.H., Genner, M.J., Hudson, A.G., Partridge, J.C., Wagner, H.-J., 2016. Localisation and origin of the bacteriochlorophyll-derived photosensitizer in the retina of the deep-sea dragon fish *Malacosteus niger*. Sci. Rep. 6, 39395. https://doi.org/10.1038/srep39395

Douglas, R.H., Partridge, J.C., Dulai, K., Hunt, D., Mullineaux, C.W., Tauber, A.Y., Hynninen, P.H., 1998a. Dragon fish see using chlorophyll. Nature 393, 423–424.

Douglas, R.H., Partridge, J.C., Dulai, K.S., Hunt, D.M., Mullineaux, C.W., Hynninen, P.H., 1999. Enhanced retinal longwave sensitivity using a chlorophyll-derived photosensitiser in Malacosteus niger, a deep-sea dragon fish with far red bioluminescence. Vision Res. 39, 2817–2832. https://doi.org/10.1016/S0042-6989(98)00332-0

Douglas, R.H., Partridge, J.C., Marshall, N.J., 1998b. The eyes of deep-sea fish. I: Lens pigmentation, tapeta and visual pigments. Prog Retin Eye Res 17, 597–636. https://doi.org/10.1016/S1350-9462(98)00002-0

Douglas, R.H., Thorpe, A., 1992. Short-wave absorbing pigments in the ocular lenses of deep-sea teleosts. J. Mar. Biol. Assoc. U. K. 72, 93–112. https://doi.org/10.1017/S0025315400048815

Doukas, A.G., Aton, B., Callender, R.H., Ebrey, T.G., 1978. Resonance Raman studies of bovine metarhodopsin I and metarhodopsin II. Biochemistry (Mosc.) 17, 2430–2435. https://doi.org/10.1021/bi00605a028

Dowling, J.E., 1999. Retinal processing of visual information. Brain Res. Bull. 50, 317. https://doi.org/10.1016/S0361-9230(99)00154-9

Dragon fish see using chlorophyll | Nature [WWW Document], n.d. URL https://www.nature.com/articles/30871 (accessed 1.10.18).

Dray, S., Josse, J., 2015. Principal component analysis with missing values: a comparative survey of methods. Plant Ecol. 216, 657–667. https://doi.org/10.1007/s11258-014-0406-z

Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A., Berson, E.L., 1990. A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature 343, 364. https://doi.org/10.1038/343364a0

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H., Calos, M.P., 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol. Cell. Biol. 7, 379–387. https://doi.org/10.1128/MCB.7.1.379

Dutta, A., Kim, T.-Y., Moeller, M., Wu, J., Alexiev, U., Klein-Seetharaman, J., 2010a. Characterization of Membrane Protein Non-native States. 2. The SDS-Unfolded States of Rhodopsin. Biochemistry (Mosc.) 49, 6329–6340. https://doi.org/10.1021/bi100339x

Dutta, A., Tirupula, K.C., Alexiev, U., Klein-Seetharaman, J., 2010b. Characterization of Membrane Protein Non-native States. 1. Extent of Unfolding and Aggregation of Rhodopsin in the Presence of Chemical Denaturants. Biochemistry (Mosc.) 49, 6317–6328. https://doi.org/10.1021/bi100338e

Elzhov, T.V., Mullen, K.M., Spiess, A.-N., Bolker, B., 2016. minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds.

Engelman, D.M., Chen, Y., Chin, C.-N., Curran, A.R., Dixon, A.M., Dupuy, A.D., Lee, A.S., Lehnert, U., Matthews, E.E., Reshetnyak, Y.K., Senes, A., Popot, J.-L., 2003. Membrane protein folding: beyond the two stage model. FEBS Lett., 126th Nobel Symposium. Membrane Proteins: Structure, Function and Assembly 555, 122–125. https://doi.org/10.1016/S0014-5793(03)01106-2

Etzkorn, M., Raschle, T., Hagn, F., Gelev, V., Rice, A.J., Walz, T., Wagner, G., 2013. Cell-free Expressed Bacteriorhodopsin in Different Soluble Membrane Mimetics: Biophysical Properties and NMR Accessibility. Structure 21, 394–401. https://doi.org/10.1016/j.str.2013.01.005

Fersht, A.R., 2000. Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. Proc. Natl. Acad. Sci. 97, 1525–1529. https://doi.org/10.1073/pnas.97.4.1525

Flötenmeyer Matthias, Weiss Hanns, Tribet Christophe, Popot Jean-Luc, Leonard Kevin, 2007. The use of amphipathic polymers for cryo electron microscopy

of NADH:ubiquinone oxidoreductase (complex I). J. Microsc. 227, 229–235. https://doi.org/10.1111/j.1365-2818.2007.01805.x

Foster, M.P., McElroy, C.A., Amero, C.D., 2007. Solution NMR of large molecules and assemblies. Biochemistry (Mosc.) 46, 331–340. https://doi.org/10.1021/bi0621314

Franke, R.R., Sakmar, T.P., Oprian, D.D., Khorana, H.G., 1988. A single amino acid substitution in rhodopsin (lysine 248----leucine) prevents activation of transducin. J. Biol. Chem. 263, 2119–2122.

Gafvelin, G., von Heijne, G., 1994. Topological "frustration" in multispanning E. coli inner membrane proteins. Cell 77, 401–412. https://doi.org/10.1016/0092-8674(94)90155-4

Gao, Y., Westfield, G., Erickson, J.W., Cerione, R.A., Skiniotis, G., Ramachandran, S., 2017. Isolation and structure–function characterization of a signaling-active rhodopsin–G protein complex. J. Biol. Chem. 292, 14280–14289. https://doi.org/10.1074/jbc.M117.797100

Garriga, P., Liu, X., Khorana, H.G., 1996. Structure and function in rhodopsin: correct folding and misfolding in point mutants at and in proximity to the site of the retinitis pigmentosa mutation Leu-125-->Arg in the transmembrane helix C. Proc. Natl. Acad. Sci. U. S. A. 93, 4560–4564.

Gattuso, H., Monari, A., Marazzi, M., 2017. Photophysics of chlorin e6: from one- and two-photon absorption to fluorescence and phosphorescence. RSC Adv. 7, 10992–10999. https://doi.org/10.1039/C6RA28616J

Gautier, A., Mott, H.R., Bostock, M.J., Kirkpatrick, J.P., Nietlispach, D., 2010. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. Nat. Struct. Mol. Biol. 17, 768–774. https://doi.org/10.1038/nsmb.1807

Getmanova, E., Patel, A.B., Klein-Seetharaman, J., Loewen, M.C., Reeves, P.J., Friedman, N., Sheves, M., Smith, S.O., Khorana, H.G., 2004. NMR Spectroscopy of Phosphorylated Wild-Type Rhodopsin: Mobility of the Phosphorylated C-Terminus of Rhodopsin in the Dark and upon Light Activation. Biochemistry (Mosc.) 43, 1126–1133. https://doi.org/10.1021/bi030120u

Gouterman, M., 1961. Spectra of porphyrins. J. Mol. Spectrosc. 6, 138–163. https://doi.org/10.1016/0022-2852(61)90236-3

Gower, J.C., 1971. A General Coefficient of Similarity and Some of Its Properties. Biometrics 27, 857–871. https://doi.org/10.2307/2528823

Grober, M.S., 1988. Brittle-star bioluminescence functions as an aposematic signal to deter crustacean predators. Anim. Behav. 36, 493–501. https://doi.org/10.1016/S0003-3472(88)80020-4

Grourke, M.J., Gibbs, J.H., 1967. Transition from random coil to α-helix induced by sodium dodecyl sulfate. Biopolymers 5, 586–588. https://doi.org/10.1002/bip.1967.360050613

Han, M., Lin, S.W., Minkova, M., Smith, S.O., Sakmar, T.P., 1996. Functional Interaction of Transmembrane Helices 3 and 6 in Rhodopsin REPLACEMENT OF PHENYLALANINE 261 BY ALANINE CAUSES REVERSION OF PHENOTYPE OF A GLYCINE 121 REPLACEMENT MUTANT. J. Biol. Chem. 271, 32337–32342. https://doi.org/10.1074/jbc.271.50.32337

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557–580. https://doi.org/10.1016/S0022-2836(83)80284-8

Hastie, T., Tibshirani, R., Friedman, J., 2009. Model Assessment and Selection, in: The Elements of Statistical Learning, Springer Series in Statistics. Springer, New York, NY, pp. 219–259. https://doi.org/10.1007/978-0-387-84858-7_7

Haugland, M.M., Anderson, E.A., Lovett, J.E., 2016. Tuning the properties of nitroxide spin labels for use in electron paramagnetic resonance spectroscopy through chemical modification of the nitroxide framework, in: Chechik, V., Murphy, D.M. (Eds.), Electron Paramagnetic Resonance. Royal Society of Chemistry, Cambridge, pp. 1–34. https://doi.org/10.1039/9781782629436-00001

Heck, M., Schädel, S.A., Maretzki, D., Bartl, F.J., Ritter, E., Palczewski, K., Hofmann, K.P., 2003. Signaling States of Rhodopsin. J. Biol. Chem. 278, 3162–3169. https://doi.org/10.1074/jbc.M209675200

Herring, P.J., 2007. Sex with the lights on? A review of bioluminescent sexual dimorphism in the sea. J. Mar. Biol. Assoc. U. K. 87, 829–842. https://doi.org/10.1017/S0025315407056433

Hespenheide, B.M., Rader, A.J., Thorpe, M.F., Kuhn, L.A., 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. J. Mol. Graph. Model. 21, 195–207. https://doi.org/10.1016/S1093-3263(02)00146-8

Hilger, D., Masureel, M., Kobilka, B.K., 2018. Structure and dynamics of GPCR signaling complexes. Nat. Struct. Mol. Biol. 25, 4–12. https://doi.org/10.1038/s41594-017-0011-7

Howland, H.C., Murphy, C.J., Mccosker, J.E., 1992. Detection of eyeshine by flashlight fishes of the family anomalopidae. Vision Res. 32, 765–769. https://doi.org/10.1016/0042-6989(92)90191-K

Huang, K.S., Bayley, H., Liao, M.J., London, E., Khorana, H.G., 1981. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. J. Biol. Chem. 256, 3802–3809.

Hwa, J., Reeves, P.J., Klein-Seetharaman, J., Davidson, F., Khorana, H.G., 1999. Structure and function in rhodopsin: further elucidation of the role of the intradiscal cysteines, Cys-110, -185, and -187, in rhodopsin folding and function. Proc Natl Acad Sci U A 96, 1932–5.

Iannaccone, A., Man, D., Waseem, N., Jennings, B.J., Ganapathiraju, M., Gallaher, K., Reese, E., Bhattacharya, S.S., Klein-Seetharaman, J., 2006. Retinitis pigmentosa associated with rhodopsin mutations: Correlation between phenotypic variability and molecular effects. Vision Res., Rhodopsin Special Issue 46, 4556–4567. https://doi.org/10.1016/j.visres.2006.08.018

Ilyas Washington, Celeste Brooks, Nicholas J. Turro, * and, Nakanishi*, K., 2004. Porphyrins As Photosensitizers To Enhance Night Vision.

Introduction, 2008. , in: Finding Groups in Data. Wiley-Blackwell, pp. 1–67. https://doi.org/10.1002/9780470316801.ch1

Isayama, T., Alexeev, D., Makino, C.L., Washington, I., Nakanishi, K., Turro, N.J., 2006. An accessory chromophore in red vision. Nature 443, 649. https://doi.org/10.1038/443649a

Isberg, V., Mordalski, S., Munk, C., Rataj, K., Harpsøe, K., Hauser, A.S., Vroling, B., Bojarski, A.J., Vriend, G., Gloriam, D.E., 2016. GPCRdb: an information system for G protein-coupled receptors. Nucleic Acids Res. 44, D356–D364. https://doi.org/10.1093/nar/gkv1178

Ishima, R., Torchia, D.A., 2000. Protein dynamics from NMR. Nat. Struct. Mol. Biol. 7, 740–743. https://doi.org/10.1038/78963

Jacobs, D.J., Rader, A.J., Kuhn, L.A., Thorpe, M.F., 2001. Protein flexibility predictions using graph theory. Proteins Struct. Funct. Bioinforma. 44, 150–165.

Jäger, S., Han, M., Lewis, J.W., Szundi, I., Sakmar, T.P., Kliger, D.S., 1997. Properties of Early Photolysis Intermediates of Rhodopsin Are Affected by Glycine 121 and Phenylalanine 261. Biochemistry (Mosc.) 36, 11804–11810. https://doi.org/10.1021/bi971122f

Jares-Erijman, E.A., Jovin, T.M., 2003. FRET imaging. Nat. Biotechnol. 21, 1387–1395. https://doi.org/10.1038/nbt896

Jastrzebska, B., Ringler, P., Palczewski, K., Engel, A., 2013. THE RHODOPSIN-TRANSDUCIN COMPLEX HOUSES TWO DISTINCT RHODOPSIN MOLECULES. J. Struct. Biol. 182, 164–172. https://doi.org/10.1016/j.jsb.2013.02.014

Jeon, C.-J., Strettoi, E., Masland, R.H., 1998. The Major Cell Populations of the Mouse Retina. J. Neurosci. 18, 8936–8946. https://doi.org/10.1523/JNEUROSCI.18-21-08936.1998

Johnsen, S., 2005. The Red and the Black: Bioluminescence and the Color of Animals in the Deep Sea. Integr. Comp. Biol. 45, 234–246. https://doi.org/10.1093/icb/45.2.234

Johnsen, S., Widder, E.A., Mobley, C.D., 2004. Propagation and Perception of Bioluminescence: Factors Affecting Counterillumination as a Cryptic Strategy. Biol. Bull. 207, 1–16. https://doi.org/10.2307/1543624

Judith Klein-Seetharaman, §, John Hwa, §, Kewen Cai, §, Christian Altenbach, ‖, Wayne L. Hubbell, ‖ and, H. Gobind Khorana*, §, 1999. Single-Cysteine Substitution Mutants at Amino Acid Positions 55−75, the Sequence Connecting the Cytoplasmic Ends of Helices I and II in Rhodopsin: Reactivity of the Sulfhydryl Groups and Their Derivatives Identifies a Tertiary Structure that Changes upon Light-Activation†,‡.

Junge, F., Schneider, B., Reckel, S., Schwarz, D., Dötsch, V., Bernhard, F., 2008. Large-scale production of functional membrane proteins. Cell. Mol. Life Sci. 65, 1729–1755. https://doi.org/10.1007/s00018-008-8067-5

Kane, R.S., Deschatelets, P., Whitesides, G.M., 2003. Kosmotropes Form the Basis of Protein-Resistant Surfaces. Langmuir 19, 2388–2391. https://doi.org/10.1021/la020737x

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Kaufman, R.J., 1985. Identification of the components necessary for adenovirus translational control and their utilization in cDNA expression vectors. Proc. Natl. Acad. Sci. U. S. A. 82, 689–693.

Kaushal, S., Khorana, H.G., 1994. Structure and Function in Rhodopsin. 7. Point Mutations Associated with Autosomal Dominant Retinitis Pigmentosa. Biochemistry (Mosc.) 33, 6121–6128. https://doi.org/10.1021/bi00186a011

Kaushal, S., Ridge, K.D., Khorana, H.G., 1994. Structure and function in rhodopsin: the role of asparagine-linked glycosylation. Proc. Natl. Acad. Sci. U. S. A. 91, 4024–4028.

Kim, D.N., Sanbonmatsu, K.Y., 2017. Tools for the cryo-EM gold rush: going from the cryo-EM map to the atomistic model. Biosci. Rep. 37. https://doi.org/10.1042/BSR20170072

Kim, M.K., Kang, Y.K., 1999. Positional preference of proline in alpha-helices. Protein Sci. Publ. Protein Soc. 8, 1492–1499.

Kirschfeld, K., Vogt, K., 1986. Does retinol serve a sensitizing function in insect photoreceptors? Vision Res. 26, 1771–1777. https://doi.org/10.1016/0042-6989(86)90127-6

Kjaergaard, M., 2015. Can proteins be intrinsically disordered inside a membrane? Intrinsically Disord. Proteins 3. https://doi.org/10.4161/21690707.2014.984570

Klein-Seetharaman, J., 2005. Dual role of interactions between membranous and soluble portions of helical membrane receptors for folding and signaling. Trends Pharmacol. Sci. 26, 183–189. https://doi.org/10.1016/j.tips.2005.02.009

Klein-Seetharaman, J., 2002. Dynamics in Rhodopsin. ChemBioChem 3, 981–986. https://doi.org/10.1002/1439-7633(20021004)3:10<981::AID-CBIC981>3.0.CO;2-9

Klein-Seetharaman, J., Getmanova, E.V., Loewen, M.C., Reeves, P.J., Khorana, H.G., 1999. NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution 19F NMR. Proc. Natl. Acad. Sci. 96, 13744–13749. https://doi.org/10.1073/pnas.96.24.13744

Klein-Seetharaman, Judith, Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M., Schwalbe, H., 2002. Long-Range Interactions Within a Nonnative Protein. Science 295, 1719–1722. https://doi.org/10.1126/science.1067680

Klein-Seetharaman, J., Reeves, P.J., Loewen, M.C., Getmanova, E.V., Chung, J., Schwalbe, H., Wright, P.E., Khorana, H.G., 2002. Solution NMR spectroscopy of [α-15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. Proc. Natl. Acad. Sci. 99, 3452–3457. https://doi.org/10.1073/pnas.052713999

Klein-Seetharaman, J., Yanamala, N.V.K., Javeed, F., Reeves, P.J., Getmanova, E.V., Loewen, M.C., Schwalbe, H., Khorana, H.G., 2004. Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. Proc. Natl. Acad. Sci. 101, 3409–3413. https://doi.org/10.1073/pnas.0308713101

Konuma Tsuyoshi, Lee Young-Ho, Goto Yuji, Sakurai Kazumasa, 2012. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. Proteins Struct. Funct. Bioinforma. 81, 107–118. https://doi.org/10.1002/prot.24166

Krebs, M.P., Holden, D.C., Joshi, P., Clark, C.L., Lee, A.H., Kaushal, S., 2010. Molecular Mechanisms of Rhodopsin Retinitis Pigmentosa and the Efficacy of Pharmacological Rescue. J. Mol. Biol. 395, 1063–1078. https://doi.org/10.1016/j.jmb.2009.11.015

Kruh, J., 1981. Effects of sodium butyrate, a new pharmacological agent, on cells in culture. Mol. Cell. Biochem. 42, 65–82. https://doi.org/10.1007/BF00222695

Krumm, B.E., Grisshammer, R., 2015. Peptide ligand recognition by G protein-coupled receptors. Front. Pharmacol. 6. https://doi.org/10.3389/fphar.2015.00048

Kvíčalová, Z., Alster, J., Hofmann, E., Khoroshyy, P., Litvín, R., Bína, D., Polívka, T., Pšenčík, J., 2016. Triplet–triplet energy transfer from chlorophylls to carotenoids in two antenna complexes from dinoflagellate Amphidinium carterae. Biochim. Biophys. Acta BBA - Bioenerg. 1857, 341–349. https://doi.org/10.1016/j.bbabio.2016.01.008

Laemmli, U.K., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227, 680–685. https://doi.org/10.1038/227680a0

Land, M.F., Fernald, R.D., 1992. The Evolution of Eyes. Annu. Rev. Neurosci. 15, 1–29. https://doi.org/10.1146/annurev.ne.15.030192.000245

Latorraca, N.R., Venkatakrishnan, A.J., Dror, R.O., 2017. GPCR Dynamics: Structures in Motion. Chem. Rev. 117, 139–155. https://doi.org/10.1021/acs.chemrev.6b00177

Leticia U. Colmenares, Walter P. Niemczura, Alfred E. Asato, and, Liu*, R.S.H., 1996. A 19F NMR Study of Rhodopsin Analogs: Use of Vinylfluororetinal Chromophores.

Lewis, J.W., van Kuijk, F.J.G.M., Carruthers, J.A., Kliger, D.S., 1997. Metarhodopsin III Formation and Decay Kinetics: Comparison of Bovine and Human Rhodopsin. Vision Res. 37, 1–8. https://doi.org/10.1016/S0042-6989(96)00138-1

Li, R., Woodward, C., 1999. The hydrogen exchange core and protein folding. Protein Sci. 8, 1571–1590. https://doi.org/10.1110/ps.8.8.1571

Liao, M.-J., London, E., Khorana, H.G., 1983. Regeneration of the native bacteriorhodopsin structure from two chymotryptic fragments. J. Biol. Chem. 258, 9949–9955.

Liu, X., Garriga, P., Khorana, H.G., 1996. Structure and function in rhodopsin: correct folding and misfolding in two point mutants in the intradiscal domain of rhodopsin identified in retinitis pigmentosa. Proc. Natl. Acad. Sci. U. S. A. 93, 4554–4559.

Lodish, H., 2008. G Protein-coupled receptors that regulate ion channels, in: Molecular Cell Biology. pp. 640–645.

Loewen, M.C., Klein-Seetharaman, J., Getmanova, E.V., Reeves, P.J., Schwalbe, H., Khorana, H.G., 2001. Solution 19F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. Proc. Natl. Acad. Sci. 98, 4888–4892. https://doi.org/10.1073/pnas.051633098

Luo, D.-G., Yue, W.W.S., Ala-Laurila, P., Yau, K.-W., 2011. Activation of Visual Pigments by Light and Heat. Science 332, 1307–1312. https://doi.org/10.1126/science.1200172

Maier, A., 2013. Neuroscience: The Cortical Layering of Visual Processing. Curr. Biol. 23, R959–R961. https://doi.org/10.1016/j.cub.2013.09.010

Malnic, B., Godfrey, P.A., Buck, L.B., 2004. The human olfactory receptor gene family. Proc. Natl. Acad. Sci. 101, 2584–2589. https://doi.org/10.1073/pnas.0307882100

Mandal, P.K., Majumdar, A., 2004. A comprehensive discussion of HSQC and HMQC pulse sequences. Concepts Magn. Reson. 20A, 1–23. https://doi.org/10.1002/cmr.a.10095

Manglik, A., Kobilka, B., 2014. The role of protein dynamics in GPCR function: insights from the beta2AR and rhodopsin. Curr Opin Cell Biol 27, 136–43. https://doi.org/10.1016/j.ceb.2014.01.008

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11, 431–441.

Merriman, M., 1877. On the History of the Method of Least Squares. The Analyst 4, 33–36. https://doi.org/10.2307/2635472

Molday, R.S., MacKenzie, D., 2002. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. Biochemistry (Mosc.) 22, 653–660. https://doi.org/10.1021/bi00272a020

Nakamichi, H., Okada, T., 2006a. Crystallographic Analysis of Primary Visual Photochemistry. Angew. Chem. Int. Ed. 45, 4270–4273. https://doi.org/10.1002/anie.200600595

Nakamichi, H., Okada, T., 2006b. Local peptide movement in the photoreaction intermediate of rhodopsin. Proc. Natl. Acad. Sci. 103, 12729–12734. https://doi.org/10.1073/pnas.0601765103

Nilsson, D.-E., Arendt, D., 2008. Eye Evolution: The Blurry Beginning. Curr. Biol. 18, R1096–R1098. https://doi.org/10.1016/j.cub.2008.10.025

Noorwez, S.M., Malhotra, R., McDowell, J.H., Smith, K.A., Krebs, M.P., Kaushal, S., 2004. Retinoids assist the cellular folding of the autosomal dominant retinitis pigmentosa opsin mutant P23H. J Biol Chem 279, 16278–84. https://doi.org/10.1074/jbc.M312101200

Nygaard, R., Zou, Y., Dror, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A., Pan, A.C., Liu, C.W., Fung, J.J., Bokoch, M.P., Thian, F.S., Kobilka, T.S., Shaw, D.E., Mueller, L., Prosser, R.S., Kobilka, B.K., 2013. The Dynamic Process of β2-Adrenergic Receptor Activation. Cell 152, 532–542. https://doi.org/10.1016/j.cell.2013.01.008

O'Day, W.T., Fernandez, H.R., 1974. Aristostomias scintillans (Malacosteidae): a deep-sea fish with visual pigments apparently adapted to its own bioluminescence. Vis. Res 14, 545–50. https://doi.org/10.1016/0042-6989(74)90044-3

Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E.M., Shichida, Y., 2002. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. Proc. Natl. Acad. Sci. 99, 5982–5987.

Opefi, C.A., South, K., Reynolds, C.A., Smith, S.O., Reeves, P.J., 2013. Retinitis Pigmentosa Mutants Provide Insight into the Role of the N-terminal Cap in Rhodopsin Folding, Structure, and Function. J. Biol. Chem. 288, 33912–33926. https://doi.org/10.1074/jbc.M113.483032

Opella, S.J., Marassi, F.M., 2017. Applications of NMR to membrane proteins. Arch. Biochem. Biophys., Nuclear Magnetic Resonance 628, 92–101. https://doi.org/10.1016/j.abb.2017.05.011

Oprian, D.D., Molday, R.S., Kaufman, R.J., Khorana, H.G., 1987. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. Proc. Natl. Acad. Sci. U. S. A. 84, 8874–8878.

Oseroff, A.R., Ohuoha, D., Hasan, T., Bommer, J.C., Yarmush, M.L., 1986. Antibody-targeted photolysis: selective photodestruction of human T-cell leukemia cells using monoclonal antibody-chlorin e6 conjugates. Proc. Natl. Acad. Sci. 83, 8744–8748. https://doi.org/10.1073/pnas.83.22.8744

Otzen, D.E., 2002. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. Biophys J 83, 2219–30. https://doi.org/10.1016/S0006-3495(02)73982-9

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M., 2000. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science 289, 739–745. https://doi.org/10.1126/science.289.5480.739

Photodynamic Therapy for Cancer [WWW Document], n.d. . Natl. Cancer Inst. URL https://www.cancer.gov/about-cancer/treatment/types/surgery/photodynamic-fact-sheet (accessed 2.22.18).

Pointer, M.A., Carvalho, L.S., Cowing, J.A., Bowmaker, J.K., Hunt, D.M., 2007. The visual pigments of a deep-sea teleost, the pearl eye Scopelarchus analis. J. Exp. Biol. 210, 2829–2835. https://doi.org/10.1242/jeb.006064

Popot, J.-L., Berry, E.A., Charvolin, D., Creuzenet, C., Ebel, C., Engelman, D.M., Flötenmeyer, M., Giusti, F., Gohon, Y., Hervé, P., Hong, Q., Lakey, J.H., Leonard, K., Shuman, H.A., Timmins, P., Warschawski, D.E., Zito, F., Zoonens, M., Pucci, B., Tribet, C., 2003. Amphipols: polymeric surfactants for membrane biology research. Cell. Mol. Life Sci. CMLS 60, 1559–1574. https://doi.org/10.1007/s00018-003-3169-6

Popot, J.-L., Engelman, D.M., 1990. Membrane protein folding and oligomerization: the two-stage model. Biochemistry (Mosc.) 29, 4031–4037.

Popot, J.-L., Gerchman, S.-E., Engelman, D.M., 1987. Refolding of bacteriorhodopsin in lipid bilayers: a thermodynamically controlled two-stage process. J. Mol. Biol. 198, 655–676.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rader, A.J., Bahar, I., 2004. Folding core predictions from network models of proteins. Polymer, Conformational Protein Conformations 45, 659–668. https://doi.org/10.1016/j.polymer.2003.10.080

Rader, A.J., Hespenheide, B.M., Kuhn, L.A., Thorpe, M.F., 2002. Protein unfolding: Rigidity lost. Proc. Natl. Acad. Sci. U. S. A. 99, 3540–3545. https://doi.org/10.1073/pnas.062492699

Ramon, E., Marron, J., del Valle, L., Bosch, L., Andrés, A., Manyosa, J., Garriga, P., 2003. Effect of dodecyl maltoside detergent on rhodopsin stability and function. Vision Res., The retinoid cycle and retina disease 43, 3055–3061. https://doi.org/10.1016/j.visres.2003.08.009

Rees, J.F., Wergifosse, B. de, Noiset, O., Dubuisson, M., Janssens, B., Thompson, E.M., 1998. The origins of marine bioluminescence: turning oxygen defence mechanisms into deep-sea communication tools. J. Exp. Biol. 201, 1211–1221.

Reeves, P.J., Kim, J.-M., Khorana, H.G., 2002. Structure and function in rhodopsin: A tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. Proc. Natl. Acad. Sci. 99, 13413–13418. https://doi.org/10.1073/pnas.212519199

Ren, L., Martin, C.H., Wise, K.J., Gillespie, N.B., Luecke, H., Lanyi, J.K., Spudich, J.L., Birge, R.R., 2001. Molecular Mechanism of Spectral Tuning in Sensory Rhodopsin II. Biochemistry (Mosc.) 40, 13906–13914. https://doi.org/10.1021/bi0116487

Resek, J.F., Farahbakhsh, Z.T., Hubbell, W.L., Khorana, H.G., 1993. Formation of the meta II photointermediate is accompanied by conformational changes in the cytoplasmic surface of rhodopsin. Biochemistry (Mosc.) 32, 12025–12032. https://doi.org/10.1021/bi00096a012

RetNet: Disease Table [WWW Document], n.d. URL https://sph.uth.edu/RetNet/disease.htm (accessed 2.24.18).

Ridge, K.D., Lee, S.S., Yao, L.L., 1995. In vivo assembly of rhodopsin from expressed polypeptide fragments. Proc. Natl. Acad. Sci. U. S. A. 92, 3204–3208.

Ridge, K.D., Palczewski, K., 2007. Visual Rhodopsin Sees the Light: Structure and Mechanism of G Protein Signaling. J. Biol. Chem. 282, 9297–9301. https://doi.org/10.1074/jbc.R600032200

Rosenfeld, T., Alchalel, A., Ottolenghi, M., 1974. On the Role of the Triplet State in the Photoisomerizationof Retinal Isomers. J. Phys. Chem. 78, 6.

Rovati, G.E., Capra, V., Neubig, R.R., 2007. The Highly Conserved DRY Motif of Class A G Protein-Coupled Receptors: Beyond the Ground State. Mol. Pharmacol. 71, 959–964. https://doi.org/10.1124/mol.106.029470

Ruschak, A.M., Kay, L.E., 2010. Methyl groups as probes of supra-molecular structure, dynamics and function. J. Biomol. NMR 46, 75. https://doi.org/10.1007/s10858-009-9376-1

Sakurai, K., Goto, Y., 2007. Principal component analysis of the pH-dependent conformational transitions of bovine β-lactoglobulin monitored by heteronuclear NMR. Proc. Natl. Acad. Sci. 104, 15346–15351. https://doi.org/10.1073/pnas.0702112104

Saliba, R.S., Munro, P.M.G., Luthert, P.J., Cheetham, M.E., 2002. The cellular fate of mutant rhodopsin: quality control, degradation and aggresome formation. J. Cell Sci. 115, 2907–2918.

Sandberg, M.A., Weigel-DiFranco, C., Dryja, T.P., Berson, E.L., 1995. Clinical expression correlates with location of rhodopsin mutation in dominant retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. 36, 1934–1942.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., Overington, J.P., 2017. A comprehensive map of molecular drug targets. Nat. Rev. Drug Discov. 16, 19–34. https://doi.org/10.1038/nrd.2016.230

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089

Seddon, A.M., Curnow, P., Booth, P.J., 2004. Membrane proteins, lipids and detergents: not just a soap opera. Biochim. Biophys. Acta BBA - Biomembr., Lipid-Protein Interactions 1666, 105–117. https://doi.org/10.1016/j.bbamem.2004.04.011

Sekharan, S., Mooney, V.L., Rivalta, I., Kazmi, M.A., Neitz, M., Neitz, J., Sakmar, T.P., Yan, E.C.Y., Batista, V.S., 2013. Spectral Tuning of Ultraviolet Cone Pigments: An Interhelical Lock Mechanism. J. Am. Chem. Soc. 135, 19064–19067. https://doi.org/10.1021/ja409896y

Shelton, P.M.J., Gaten, E., Herring, P.J., 1992. Adaptations of tapeta in the eyes of mesopelagic decapod shrimps to match the oceanic irradiance distribution. J. Mar. Biol. Assoc. U. K. 72, 77–88. https://doi.org/10.1017/S0025315400048797

Shimomura, O., Inoue, S., Johnson, F.H., Haneda, Y., 1980. Widespread occurrence of coelenterazine in marine bioluminescence. Comp. Biochem. Physiol. Part B Comp. Biochem. 65, 435–437.

Shimomura, O., Teranishi, K., 2000. Light-emitters involved in the luminescence of coelenterazine. Luminescence 15, 51–58. https://doi.org/10.1002/(SICI)1522-7243(200001/02)15:1<51::AID-BIO555>3.0.CO;2-J

Shortle, D., Ackerman, M.S., 2001. Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. Science 293, 487–489. https://doi.org/10.1126/science.1060438

Siano, D.B., Metzler, D.E., 1969. Band Shapes of the Electronic Spectra of Complex Molecules. J. Chem. Phys. 51, 1856–1861. https://doi.org/10.1063/1.1672270

Singh-Rachford, T.N., Castellano, F.N., 2010. Photon upconversion based on sensitized triplet–triplet annihilation. Coord. Chem. Rev., 18th International Symposium on the Photochemistry and Photophysics of Coordination Compounds Sapporo, 2009 254, 2560–2573. https://doi.org/10.1016/j.ccr.2010.01.003

Sjodt, M., Clubb, R.T., 2017. Nitroxide Labeling of Proteins and the Determination of Paramagnetic Relaxation Derived Distance Restraints for NMR Studies. Bio-Protoc. 7. https://doi.org/10.21769/BioProtoc.2207

Smith, R.C., Baker, K.S., 1981. Optical properties of the clearest natural waters (200–800 nm). Appl. Opt. 20, 177–184. https://doi.org/10.1364/AO.20.000177

Somiya, H., 1982. `Yellow Lens' Eyes of a Stomiatoid Deep-Sea Fish, Malacosteus niger. Proc. R. Soc. Lond. B Biol. Sci. 215, 481–489.

Soong, B.-W., Huang, Y.-H., Tsai, P.-C., Huang, C.-C., Pan, H.-C., Lu, Y.-C., Chien, H.-J., Liu, T.-T., Chang, M.-H., Lin, K.-P., Tu, P.-H., Kao, L.-S., Lee, Y.-C., 2013. Exome sequencing identifies GNB4 mutations as a cause of dominant intermediate Charcot-Marie-Tooth disease. Am. J. Hum. Genet. 92, 422–430. https://doi.org/10.1016/j.ajhg.2013.01.014

Stancik, A.L., Brauns, E.B., 2008. A simple asymmetric lineshape for fitting infrared absorption spectra. Vib. Spectrosc. 47, 66–69. https://doi.org/10.1016/j.vibspec.2008.02.009

Steinberg, G., Ottolenghi, M., Sheves, M., 1993. pKa of the protonated Schiff base of bovine rhodopsin. A study with artificial pigments. Biophys. J. 64, 1499–1502.

Sternberg, S.R., 1983. Biomedical Image Processing. Computer 16, 22–34. https://doi.org/10.1109/MC.1983.1654163

Stigler, S.M., 1981. Gauss and the Invention of Least Squares. Ann. Stat. 9, 465–474. https://doi.org/10.1214/aos/1176345451

Sudhoelter, E.J.R., De Grip, W.J., Engberts, J.B.F.N., 1982. Rhodopsin reconstitution in vesicles formed from simple, fully synthetic amphiphiles. J. Am. Chem. Soc. 104, 1069–1072. https://doi.org/10.1021/ja00368a025

Sutton, T.T., 2005. Trophic ecology of the deep-sea fish Malacosteus niger (Pisces: Stomiidae): An enigmatic feeding ecology to facilitate a unique visual system? Deep Sea Res. Part Oceanogr. Res. Pap. 52, 2065–2076. https://doi.org/10.1016/j.dsr.2005.06.011

Tam, B.M., Moritz, O.L., 2009. The Role of Rhodopsin Glycosylation in Protein Folding, Trafficking, and Light-Sensitive Retinal Degeneration. J. Neurosci. 29, 15145–15154. https://doi.org/10.1523/JNEUROSCI.4259-09.2009

Tarentino, A.L., Trimble, R.B., Plummer, T.H., 1989. Chapter 5 Enzymatic Approaches for Studying the Structure, Synthesis, and Processing of Glycoproteins, in: Tartakoff, A.M. (Ed.), Methods in Cell Biology, Vesicular Transport Part B. Academic Press, pp. 111–139. https://doi.org/10.1016/S0091-679X(08)61169-3

Tastan, O., Dutta, A., Booth, P., Klein-Seetharaman, J., 2014. Retinal proteins as model systems for membrane protein folding. Biochim. Biophys. Acta BBA - Bioenerg., Retinal Proteins 1837, 656–663. https://doi.org/10.1016/j.bbabio.2013.11.021

Tastan, O., Yu, E., Ganapathiraju, M., Aref, A., Rader, A.J., Klein-Seetharaman, J., 2007. Comparison of Stability Predictions and Simulated Unfolding of Rhodopsin Structures†. Photochem. Photobiol. 83, 351–363. https://doi.org/10.1562/2006-06-20-RA-942

Teller, D.C., Stenkamp, R.E., Palczewski, K., 2003. Evolutionary analysis of rhodopsin and cone pigments: connecting the three-dimensional structure with spectral tuning and signal transfer. FEBS Lett., 126th Nobel Symposium. Membrane Proteins: Structure, Function and Assembly 555, 151–159. https://doi.org/10.1016/S0014-5793(03)01152-9

Tribet, C., Audebert, R., Popot, J.-L., 1996. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. Proc. Natl. Acad. Sci. 93, 15047–15050. https://doi.org/10.1073/pnas.93.26.15047

Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E., Kay, L.E., 2003. Cross-Correlated Relaxation Enhanced ^{1}H– ^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. J. Am. Chem. Soc. 125, 10420–10428. https://doi.org/10.1021/ja030153x

Ulloa-Aguirre, A., Zariñán, T., Dias, J.A., Conn, P.M., 2014. Mutations in G protein-coupled receptors that impact receptor trafficking and reproductive function. Mol. Cell. Endocrinol. 382. https://doi.org/10.1016/j.mce.2013.06.024

Vitrac, H., MacLean, D.M., Jayaraman, V., Bogdanov, M., Dowhan, W., 2015. Dynamic membrane protein topological switching upon changes in phospholipid environment. Proc. Natl. Acad. Sci. 112, 13874–13879. https://doi.org/10.1073/pnas.1512994112

Vogel, R., Mahalingam, M., Lüdeke, S., Huber, T., Siebert, F., Sakmar, T.P., 2008. Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. J. Mol. Biol. 380, 648–655. https://doi.org/10.1016/j.jmb.2008.05.022

Wang, D., Zhao, W.-L., Cai, M.-J., Wang, J.-X., Zhao, X.-F., 2015. G-protein-coupled receptor controls steroid hormone signaling in cell membrane. Sci. Rep. 5. https://doi.org/10.1038/srep08675

Warrant, E., 2000. The eyes of deep–sea fishes and the changing nature of visual scenes with depth. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 1155–1159. https://doi.org/10.1098/rstb.2000.0658

Washington, I., Zhou, J., Jockusch, S., Turro, N.J., Nakanishi, K., Sparrow, J.R., 2007. Chlorophyll derivatives as visual pigments for super vision in the red. Photochem. Photobiol. Sci. 6, 775–779. https://doi.org/10.1039/B618104J

Werner, K., Lehner, I., Dhiman, H.K., Richter, C., Glaubitz, C., Schwalbe, H., Klein-Seetharaman, J., Khorana, H.G., 2007. Combined solid state and solution NMR studies of α,ε-15N labeled bovine rhodopsin. J. Biomol. NMR 37, 303–312. https://doi.org/10.1007/s10858-007-9143-0

Werner, K., Richter, C., Klein-Seetharaman, J., Schwalbe, H., 2008. Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. J. Biomol. NMR 40, 49–53. https://doi.org/10.1007/s10858-007-9205-3

Widder, E.A., Latz, M.I., Herring, P.J., Case, J.F., 1984. Far red bioluminescence from two deep-sea fishes. Science 225, 512–4. https://doi.org/10.1126/science.225.4661.512

Wishart, D.S., Sykes, S.B.D., Richards, F.M., 1992. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry (Mosc.) 1647–1651.

Wong, G.G., Witek, J.S., Temple, P.A., Wilkens, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A., Clark, S.C., 1985. Human GM-CSF: Molecular Cloning of the Complementary DNA and Purification of the Natural and Recombinant Proteins. Science 228, 810–815.

Woodall, N.B., Yin, Y., Bowie, J.U., 2015. Dual-topology insertion of a dual-topology membrane protein. Nat. Commun. 6. https://doi.org/10.1038/ncomms9099

Woods, K.N., 2014. Using THz time-scale infrared spectroscopy to examine the role of collective, thermal fluctuations in the formation of myoglobin allosteric communication pathways and ligand specificity. https://doi.org/10.1039/C3SM53229A

Woods, K.N., Pfeffer, J., Klein-Seetharaman, J., 2017. Chlorophyll-Derivative Modulation of Rhodopsin Signaling Properties through Evolutionarily Conserved Interaction Pathways. Front. Mol. Biosci. 4. https://doi.org/10.3389/fmolb.2017.00085

Ye, S., Zaitseva, E., Caltabiano, G., Schertler, G.F.X., Sakmar, T.P., Deupi, X., Vogel, R., 2010. Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. Nature 464, 1386–1389. https://doi.org/10.1038/nature08948

Yokoyama, R., Knox, B.E., Yokoyama, S., 1995. Rhodopsin from the fish, Astyanax: role of tyrosine 261 in the red shift. Invest. Ophthalmol. Vis. Sci. 36, 939–945.

Yokoyama, S., Tada, T., Yamato, T., 2007. Modulation of the Absorption Maximum of Rhodopsin by Amino Acids in the C-terminus. Photochem. Photobiol. 83, 236–241. https://doi.org/10.1562/2006-06-19-RA-939

Zenkevich, E., Sagun, E., Knyukshto, V., Shulga, A., Mironov, A., Efremova, O., Bonnett, R., Songca, S.P., Kassem, M., 1996. Photophysical and photochemical properties of potential porphyrin and chlorin photosensitizers for PDT. J. Photochem. Photobiol. B 33, 171–180. https://doi.org/10.1016/1011-1344(95)07241-1